

# Representer Theorems in Banach Spaces: Minimum Norm Interpolation, Regularized Learning and Semi-Discrete Inverse Problems

**Rui Wang**

*School of Mathematics  
Jilin University  
Changchun, 130012, P. R. China*

RWANG11@JLU.EDU.CN

**Yuesheng Xu**

*Department of Mathematics and Statistics  
Old Dominion University  
Norfolk, VA 23529, USA*

Y1XU@ODU.EDU

**Editor:** John Shawe-Taylor

## Abstract

Learning a function from a finite number of sampled data points (measurements) is a fundamental problem in science and engineering. This is often formulated as a minimum norm interpolation (MNI) problem, a regularized learning problem or, in general, a semi-discrete inverse problem (SDIP), in either Hilbert spaces or Banach spaces. The goal of this paper is to systematically study solutions of these problems in Banach spaces. We aim at obtaining *explicit* representer theorems for their solutions, on which convenient solution methods can then be developed. For the MNI problem, the explicit representer theorems enable us to express the infimum in terms of the norm of the linear combination of the interpolation functionals. For the purpose of developing efficient computational algorithms, we establish the fixed-point equation formulation of solutions of these problems. We reveal that unlike in a Hilbert space, in general, solutions of these problems in a Banach space may not be able to be reduced to *truly* finite dimensional problems (with certain infinite dimensional components hidden). We demonstrate how this obstacle can be removed, reducing the original problem to a truly finite dimensional one, in the special case when the Banach space is  $\ell_1(\mathbb{N})$ .

**Keywords:** representer theorem, minimum norm interpolation, regularized learning, sparse learning, semi-discrete inverse problem, Banach space

## 1. Introduction

A core issue in data science is to learn a function from a finite number of sampled data points. This may be modeled as an interpolation problem, an optimization problem or, a semi-discrete inverse problem (SDIP). Learning such a function is an ill-posed problem in the sense that a small error in sampled data may result in a large error in the resulting function. Because sampled data inevitably contain noise, the ill-posedness of these problems is unavoidable. It is well-recognized that minimum norm interpolation (MNI) and the regularization method are effective approaches to treat the ill-posedness. The goal of this paper is to systematically study the solution representation of the three types of problems: MNI,

regularized learning and regularized SDIPs in Banach spaces. Regularized learning and regularized SDIPs are originated from different sources. The SDIP often refers to a physical problem described by a physical law expressed via a certain integral equation, which relates its solution with a finite number of measurements. While a regularized learning problem has a fidelity term describing a certain learning approach such as a network (not necessarily a physical law). However, these two types of problems have the same general mathematical formulation in the sense that both of them have a fidelity term involving a finite number of measurements and a regularization term specifying prior solution information. For this reason, we will not distinguish them in this paper since as far as the solution representation is concerned, there is little distinct between them. For simplicity we may use the term the “regularization problem” to refer to both of these problems, when necessary.

Classical regularization methods aim at finding a target function in a reproducing kernel Hilbert space (RKHS) on which point-evaluation functionals are continuous, from a finite number of point-evaluation functionals. The point-evaluation functionals on an RKHS can be represented by the reproducing kernel (Aronszajn, 1950). Applications of kernels in machine learning were reviewed by Cucker and Smale (2002). Connections of reproducing kernels with sampling were discussed by Smale and Zhou (2004). Recent applications of kernel functions in numerical solutions of partial differential equations may be found in Schaback and Wendland (2006). Reproducing kernels and associated RKHSs over infinite, discrete and countable sets were studied by Jorgensen and Tian (2015). The success of kernel based regularization methods lies on the celebrated representer theorem (Argyriou et al., 2009; Cox and O’Sullivan, 1990; Kimeldorf and Wahba, 1970; Schölkopf et al., 2001), which states that a solution of the regularization problem is a linear combination of kernel sessions (the kernel with one of its variable evaluated at given data points). The earliest form of the representer theorem in a Hilbert space may be traced back to de Boor and Lynch (1966) and Kimeldorf and Wahba (1970). The representer theorem of Kimeldorf and Wahba has been found applicable to the solution of the SDIP (Krebs et al., 2009; Wendland, 2005). A multivariate version of the L-spline smoothing problem was investigated by De Figueiredo and Chen (1990), giving a representer theorem for its solution. In the context of machine learning, the representer theorem for the solution of the regularized empirical risk minimization in an RKHS was established by Schölkopf, Herbrich and Smola (2001). Argyriou, Micchelli and Pontil (2009) gave necessary and sufficient conditions to ensure that a general regularized empirical risk minimization problem in an RKHS has a representer theorem. Moreover, we obtained in our recent work (Wang and Xu, 2019, 2021) representer theorems for solutions of the regularized SDIPs in the functional RKHSs naturally introduced by the inverse problems.

The representer theorem is useful in both theory and computation. Representer theorems for the regularization problem reveal exactly in what sub-class its solution lies. According to the representer theorem, a solution of the problem is a linear combination of the kernel sessions. This leads to the study of universality of a kernel by Micchelli, Xu and Zhang (2006), which gives necessary and sufficient conditions on a kernel so that a linear combination of the kernel sessions can arbitrarily approximate a given continuous function. Moreover, motivated by representing a solution of the regularized learning in a multiscale manner, refinement of a reproducing kernel was studied in (Xu and Zhang, 2007, 2009; Zhang et al., 2012). From a practical standpoint, representer theorems for either MNI or

regularization problems are useful, because they dramatically reduce an infinite dimensional problem to a finite dimensional one whose solution can be obtained by solving either a linear system or a finite dimensional optimization problem.

Compared to Hilbert spaces, Banach spaces with more choices of norms enjoy more geometric structures, some of which can promote sparsity for learning solutions in these spaces. Several recent research directions point to consideration of MNI, regularized learning or SDIPs in Banach spaces. Compress sensing (Candès et al., 2006; Donoho, 2006) motivates us to study MNI or the regularized learning problem in Banach spaces. Image restoration using TV norms for regularization (Cai et al., 2012; Rudin et al., 1992) leads to searching an optimization solution in a Banach space. SDIPs were recently considered in Banach spaces (Schuster et al., 2012). Regularized learning in Banach spaces was originated in Micchelli and Pontil (2004). Since then, regularized learning in a Banach space and a desired representer theorem of its solutions have received considerable attention in the literature. The MNI or its related regularization problem in a Banach space is also motivated from a theoretical point of view: the functional extension problem in such a space. Recently, extension of a given function on a finite set in  $\mathbb{R}^n$  to a function on the entire  $\mathbb{R}^n$  was studied in a series of papers (Fefferman, 2005, 2009; Fefferman and B. Klartag, 2009a,b).

The notion of reproducing kernel Banach space (RKBS) was originally introduced in Zhang et al. (2009) and further developed in Sriperumbudur et al. (2011); Song and Zhang (2011); Song et al. (2013); Xu and Ye (2019); Zhang and Zhang (2012). In the framework of a semi-inner-product RKBS, the representer theorem of the solutions of the regularized learning problem was derived from the dual elements and the semi-inner-product (Zhang et al., 2009; Zhang and Zhang, 2012). In Xu and Ye (2019), an alternative definition of RKBS was provided by the dual bilinear form. In that paper, for a reflexive and smooth RKBS, the representer theorem of the solutions of the regularized learning problem was also obtained using the Gâteaux derivative of the norm function and the reproducing kernel. The above RKBSs, in which the representer theorem was well established for the regularized learning problem, are all reflexive and smooth. In fact, the reflexivity guarantees the existence of solutions of the regularized learning problem and the smoothness allows us to use the Gâteaux derivative of the norm function to describe the representer theorem. In the special case of a semi-inner-product RKBS, the Gâteaux derivative can be represented by the semi-inner-product. In addition, the reproducing kernel provides a closed-form function representation for the point-evaluation functionals. The representer theorem was generalized to a non-reflexive and non-smooth Banach space which has a pre-dual space (Huang et al., 2020; Unser, 2016, 2019b). Having a pre-dual space guarantees that the Banach space has the weak\* topology, which together with the continuity of the loss function and the regularizer, also leads to the existence of the solutions. Due to lack of the Gâteaux derivative, other tools need to be used to describe the representer theorem. The representer theorem was obtained in Huang et al. (2020) by employing the subdifferential of the norm function for a lower semi-continuous loss function and the quadratic regularizer. Representer theorems for a class of inverse problems with a convex and continuous loss function and regularizer was established in Unser (2019b) by the duality mapping. Moreover, representer theorems for deep kernel learning and deep neural networks were obtained in Bohn (2019) and Unser (2019a), respectively.

It is the main purpose of this paper to understand the solution representation of the MNI problem and the regularization problem in a Banach space. In the literature there are a few existing representer theorems for regularized learning problems in a Banach space. However, all of them are in *implicit* forms, which may not be convenient for direct solution representations. We will first bridge different approaches used in the literature for establishing representer theorems for a solution of the MNI problem in a Banach space and its related regularization problem, to deepen the understanding of the underlying functional analytic ideas. As such, we will provide novel *explicit* representer theorems in a general setting, potentially useful for direct solution representations. We aim at revealing the *simplicity, beauty, generality and unity* of the representer theorem and commit to developing solution representations of these problems suitable for further design of numerical algorithms. Major mathematical contributions made in this paper include the following five aspects:

- We have conducted a systematic study of the representer theorems for a solution of the MNI problem and the regularization problem in a Banach space, by using both functional analytic and convex analytic approaches.
- We have established explicit solution representations for the MNI problem and the regularization problem in a Banach space which has a dual space in both smooth and non-smooth cases.
- We have developed approaches to determine the coefficients appearing in solution representations of these problems, which can be determined by solving a linear/nonlinear system, or a finite dimensional optimization problem, leading to solution methods for solving them.
- We have expressed the infimum of the MNI in a Banach space in terms of the interpolation functionals, by using its solution representations and properties of the subdifferential of the norm function of the Banach space.
- We have observed that although the representer theorem in a Banach space converts the originally infinite dimensional problem to a finite dimensional one, unlike in a Hilbert space where the resulting linear system is truly finite dimensional, the resulting finite dimensional problem has certain hidden infinite dimensional components, and we have demonstrated a way to overcome this challenge in the special Banach space  $\ell_1(\mathbb{N})$ .

The MNI problem is closely related to the regularized learning problem, see Micchelli and Pontil (2004). We first establish solution representations for the MNI problem, and then convert the resulting representer theorems to the regularization problem through the relation between the two problems. The essence of the representer theorem refers to that the original optimization problem in an infinite dimensional space can be reduced to one possibly in a finite dimensional space. This profits from the fact that the number of data points used to learn a function is finite. A crucial issue about the representer theorem concerns how to characterize the relation between the solutions of the original infinite dimensional optimization problem and the finite dimensional one. To address it, we characterize the MNI problem through two different approaches. Firstly, the MNI problem is interpreted

as a problem of best approximation. Due to the Hahn-Banach theorem, the latter can be characterized by the functionals which peak at the best approximation point. The set of such functionals is defined by the value of the duality mapping at the point. Accordingly, the duality mapping becomes a suitable tool for the representer theorem of the solutions of the MNI problem and then the regularization problem. Secondly, as a classical optimization problem with constraints, the MNI problem can be solved by the Lagrange multiplier method. Due to lack of the smoothness of general Banach spaces (not necessarily smooth), the subdifferential of the norm function needs to be used here. In a special case that the Banach space is smooth, the duality mapping and the subdifferential of the norm function are both reduced to the Gâteaux derivative of the norm function. The representer theorem for this case has a simple form which is described by the Gâteaux derivative. In summary, the fact that the number of data used to learn a function is finite leads to the desired representer theorems and the mathematical tools, such as the duality mapping, the subdifferential and the Gâteaux derivative of the norm function help us describe the representer theorems.

It is desirable to develop solution representations of the MNI problem and the regularization problem convenient for algorithmic development. Inspired by the success of the fixed-point approach used in solving several types of finite dimensional problems such as machine learning (Li et al., 2020; Li et al., 2018, 2019; Polson et al., 2015), image processing (Chen et al., 2013; Li et al., 2012; Li et al., 2015; Lu et al., 2016; Micchelli et al., 2011), medical imaging (Krol et al., 2012; Li et al., 2015; Zheng et al., 2019) and solutions of SDIPs (Fan et al., 2014; Jin and Lu, 2014), we develop solution representations for the MNI problem and the regularization problem by using a fixed-point formulation via the proximity operator of the functions appearing in the objective function or constraints. This formulation is convenient for designing iterative algorithms for solving these problems. Difficulty of extending the existing work which is either in a finite dimensional space or in a Hilbert space to the current setting lies in the infinite dimensional component of the Banach space. In particular, we reformulate solutions of the MNI problem and the regularization problem in the special Banach space  $\ell_1(\mathbb{N})$  as fixed-points of a nonlinear map defined on a finite dimensional space by making use of special properties of a pre-dual space of  $\ell_1(\mathbb{N})$ , leading to implementable iterative algorithms for solving the problem. We remark that a solution method for the MNI problem in  $\ell_1(\mathbb{N})$  was proposed in Cheng and Xu (2020) by reformulating it as a linear programming problem. However, solving the resulting linear programming problem requires an exponential computational cost, and thus the method is not feasible for practical computation in the context of big data analytic. The fixed-point equation approach presented in this paper overcomes this difficulty.

In passing, we would like to point it out that although the representer theorem reduces an infinite dimensional problem to a finite dimensional one, in general, often certain infinite dimensional component is hidden in the resulting finite dimensional problem. We will single out these hidden infinite dimensional component and for certain special cases of practical importance, we will show how this obstacle can be removed to obtain a truly finite dimensional one. Developing efficient computational algorithms based on the solution representations provided by this paper requires further investigation. Nevertheless, the theory established here furnishes a solid mathematical foundation for this practical goal.

This paper is organized in eight sections and an appendix. In section 2, we describe the MNI problem in a Banach space and present a sufficient condition to ensure the existence

of its solutions. We characterize in section 3 a solution of the MNI problem by two different approaches in which either of the duality mapping or the subdifferential of the norm function is used to describe the representer theorem for the problem. We first establish implicit representer theorems and then derive explicit representer theorems by using duality arguments. We also consider several special cases of practical importance and provide special results for these cases. In section 4, we develop approaches to determine the coefficients involved in the representer theorems when the Banach space has a pre-dual space and the linear functionals appearing in the problem belong to the pre-dual space. These approaches of determining the coefficients lead to solution methods for solving the MNI problem. We also present the infimum of the MNI problem in a Banach space. In section 5, we propose fixed-point equations for the MNI problem in a Banach space. This fixed-point formulation will serve as a basis for further development of efficient convergence guaranteed algorithms. We describe in section 6 the regularization problem in a Banach space and propose a sufficient condition that ensures the existence of its solutions. We also elaborate an intrinsic relation between the regularization problem and a related MNI problem. In section 7, we establish both implicit and explicit representer theorems for the regularization problem. We also deliver special results for several cases of practical importance. Moreover, the second portion of section 7 is devoted to the presentation of solutions of regularization problems in Banach spaces. We present the representer theorems based solution representations and as well as the fixed-point formulation for the regularization problems. We discuss in section 8 the connection of the representer theorems established in this paper for the MNI problem and the regularization problem with the existing results in the literature. In Appendix A, we describe an example of using RKBSs in developing sparse machine learning methods, to illustrate the relevance of RKBSs to machine learning. Technical proofs for theorems of this paper are included in Appendices B-G.

## 2. MNI in a Banach Space

MNI aims at finding an element, in a suitable space, having the smallest norm and interpolating a given set of sampled data. In this section, we describe the MNI problem in a Banach space and present a sufficient condition which ensures the existence of its solutions.

We first describe the MNI problem in a Banach space. Let  $\mathcal{B}$  denote a real Banach space with norm  $\|\cdot\|_{\mathcal{B}}$ . By  $\mathcal{B}^*$  we denote the dual space of  $\mathcal{B}$ , the space of all continuous linear functionals on  $\mathcal{B}$  with the norm  $\|\nu\|_{\mathcal{B}^*} := \sup_{f \in \mathcal{B}, f \neq 0} \frac{|\nu(f)|}{\|f\|_{\mathcal{B}}}$ , for all  $\nu \in \mathcal{B}^*$ . The dual bilinear form  $\langle \cdot, \cdot \rangle_{\mathcal{B}}$  on  $\mathcal{B}^* \times \mathcal{B}$  is defined as  $\langle \nu, f \rangle_{\mathcal{B}} := \nu(f)$ , for all  $\nu \in \mathcal{B}^*$  and all  $f \in \mathcal{B}$ . For each  $m \in \mathbb{N}$ , let  $\mathbb{N}_m := \{1, 2, \dots, m\}$ . Suppose that  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , are a finite number of linearly independent elements. Associated with these functionals, we introduce an operator  $\mathcal{L} : \mathcal{B} \rightarrow \mathbb{R}^m$  by

$$\mathcal{L}(f) := [\langle \nu_j, f \rangle_{\mathcal{B}} : j \in \mathbb{N}_m], \quad \text{for all } f \in \mathcal{B}. \quad (1)$$

According to the continuity of the linear functionals  $\nu_j$ ,  $j \in \mathbb{N}_m$ , on  $\mathcal{B}$ , we have for each  $f \in \mathcal{B}$  that

$$\|\mathcal{L}(f)\|_{\mathbb{R}^m} = \left( \sum_{j \in \mathbb{N}_m} |\langle \nu_j, f \rangle_{\mathcal{B}}|^2 \right)^{1/2} \leq \left( \sum_{j \in \mathbb{N}_m} \|\nu_j\|_{\mathcal{B}^*}^2 \right)^{1/2} \|f\|_{\mathcal{B}},$$

which yields that  $\|\mathcal{L}\| \leq \left(\sum_{j \in \mathbb{N}_m} \|\nu_j\|_{\mathcal{B}^*}^2\right)^{1/2}$ . For a given vector  $\mathbf{y} := [y_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$ , we set

$$\mathcal{M}_{\mathbf{y}} := \{f \in \mathcal{B} : \mathcal{L}(f) = \mathbf{y}\}. \quad (2)$$

In particular, when  $\mathbf{y}$  is the zero vector, we write  $\mathcal{M}_0$ . The MNI problem with given data set  $\{(\nu_j, y_j) : j \in \mathbb{N}_m\}$  has the form

$$\inf\{\|f\|_{\mathcal{B}} : f \in \mathcal{M}_{\mathbf{y}}\}. \quad (3)$$

We now consider the existence of a solution of problem (3). The linear independence of the functionals  $\nu_j$ ,  $j \in \mathbb{N}_m$ , ensures that  $\mathcal{M}_{\mathbf{y}}$  is nonempty for any given  $\mathbf{y} \in \mathbb{R}^m$ . By employing standard arguments in convex analysis (Ekeland and Temam, 1999; Zălinescu, 2002), we establish a sufficient condition that ensures the existence of a solution of the problem. To this end, we recall some notions in Banach spaces. Since the natural map is the isometrically imbedding map from  $\mathcal{B}$  into  $\mathcal{B}^{**}$ , there holds

$$\langle \nu, f \rangle_{\mathcal{B}} = \langle f, \nu \rangle_{\mathcal{B}^*}, \quad \text{for all } f \in \mathcal{B} \text{ and all } \nu \in \mathcal{B}^*. \quad (4)$$

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  $\nu \rightarrow \langle \nu, f \rangle_{\mathcal{B}}$  on  $\mathcal{B}^*$  is continuous with respect to the topology. A sequence  $\nu_n$ ,  $n \in \mathbb{N}$ , in  $\mathcal{B}^*$  is said to converge weakly\* to  $\nu \in \mathcal{B}^*$  if  $\lim_{n \rightarrow +\infty} \langle \nu_n, f \rangle_{\mathcal{B}} = \langle \nu, f \rangle_{\mathcal{B}}$ , for all  $f \in \mathcal{B}$ . A normed space  $\mathcal{B}_*$  is called a pre-dual space of a Banach space  $\mathcal{B}$  if  $(\mathcal{B}_*)^* = \mathcal{B}$ . It follows from equation (4) with  $\mathcal{B}^*$  being replaced by  $\mathcal{B}_*$  that

$$\langle \nu, f \rangle_{\mathcal{B}} = \langle f, \nu \rangle_{\mathcal{B}_*}, \quad \text{for all } f \in \mathcal{B} \text{ and all } \nu \in \mathcal{B}_*. \quad (5)$$

A pre-dual space  $\mathcal{B}_*$  guarantees that the Banach space  $\mathcal{B}$  enjoys the weak\* topology. A consequence of the Banach-Alaoglu theorem (Megginson, 1998) ensures that if a Banach space  $\mathcal{B}$  has a pre-dual space  $\mathcal{B}_*$ , then any bounded and weakly\* closed subsets of  $\mathcal{B}$  is weakly\* compact. In the special case that  $\mathcal{B}$  has a separable pre-dual space  $\mathcal{B}_*$ , any bounded sequence  $f_n$ ,  $n \in \mathbb{N}$ , in  $\mathcal{B}$  has a weak\* accumulation point  $f \in \mathcal{B}$ . That is, there exists a subsequence  $f_{n_k}$ ,  $k \in \mathbb{N}$ , such that  $\lim_{k \rightarrow +\infty} \langle \nu, f_{n_k} \rangle_{\mathcal{B}} = \langle \nu, f \rangle_{\mathcal{B}}$ , for all  $\nu \in \mathcal{B}_*$ . A Banach space  $\mathcal{B}$  is said to be reflexive if  $(\mathcal{B}^*)^* = \mathcal{B}$ . It is clear that a reflexive Banach space  $\mathcal{B}$  always takes the dual space  $\mathcal{B}^*$  as a pre-dual space  $\mathcal{B}_*$ . However, a Banach space  $\mathcal{B}$  having a pre-dual space may not be reflexive. For example, the Banach space  $\ell_1(\mathbb{N})$  of all real sequences  $\mathbf{x} := (x_j : j \in \mathbb{N})$ , with  $\|\mathbf{x}\|_1 := \sum_{j \in \mathbb{N}} |x_j| < +\infty$ , has  $c_0$  as its pre-dual space, where  $c_0$  denotes the space of all real sequences  $\mathbf{u} := (u_j : j \in \mathbb{N})$  such that  $\lim_{j \rightarrow +\infty} u_j = 0$ , with  $\|\mathbf{u}\|_{\infty} := \sup\{|u_j| : j \in \mathbb{N}\} < +\infty$ . Clearly, the Banach space  $\ell_1(\mathbb{N})$  has a pre-dual space but it is not reflexive.

We now turn to considering the existence of a solution of problem (3) in a Banach space  $\mathcal{B}$  having a pre-dual space  $\mathcal{B}_*$ . In this case, the linear functionals appearing in (3) need to be restricted to the pre-dual space  $\mathcal{B}_*$ . The complete proof of the following existing result is included in Appendix B.

**Proposition 1** *If the Banach space  $\mathcal{B}$  has a pre-dual space  $\mathcal{B}_*$  and  $\nu_j \in \mathcal{B}_*$ ,  $j \in \mathbb{N}_m$ , are linearly independent, then for any  $\mathbf{y} \in \mathbb{R}^m$  problem (3) has at least one solution.*

We note that solutions of problem (3) may not be unique unless the Banach space  $\mathcal{B}$  is strictly convex. In the special case that  $\mathcal{B} = \ell_1(\mathbb{N})$ , the existence of a solution of problem (3) was given in Cheng and Xu (2020) by an elementary argument. As a consequence of the above proposition, when  $\mathcal{B}$  is reflexive, we also have the existence of a solution of problem (3), since in this case the dual space  $\mathcal{B}^*$  is its pre-dual space.

### 3. Representer Theorems for MNI

In this section, we establish several representer theorems for a solution of the MNI problem, with proofs included in Appendix C. The resulting representer theorems show that even though the MNI problem with a finite number of data points is a minimization problem in an *infinite* dimensional space, it can be transferred to one possibly in a *finite* dimensional space.

The representer theorems for the MNI problem established in the literature are often stated with restricted assumptions on the Banach space (Xu and Ye, 2019; Zhang et al., 2009). We realize that most of the assumptions were used to ensure the existence and uniqueness of the solution of the MNI problem. For example, as we have established in the last section, if  $\mathcal{B}$  is a Banach space having a pre-dual space or being reflexive, then the MNI problem has at least a solution. If  $\mathcal{B}$  is strictly convex, then there exists at most a solution of the MNI problem. The smoothness of the Banach space allows us to describe the representer theorem by using the Gâteaux derivative of the norm function. We shall clarify in this section that the validity of the representer theorem does not depend on these properties of the Banach space.

We first present implicit representer theorems for a solution of problem (3), obtained from two different approaches: functional analytic and convex analytic. In the functional analytic approach, we convert problem (3) as a best approximation problem and then use the duality theory to characterize the best approximation from a linear translate of a subspace. The duality theory was used extensively in the literature (Braess, 1986; Chui, 1990; Deutsch, 2001; Deutsch et al., 1995; Deutsch et al., 1996; Micchelli et al., 1985; Pinkus, 1989; Swetits et al., 1990a,b, 1991; Ubhaya and Xu, 1995; Xu, 1989) to characterize a best approximation from a convex set or a subspace in Banach spaces. For a nonempty subset  $\mathcal{M}$  of  $\mathcal{B}$ , we define the distance from  $f \in \mathcal{B}$  to  $\mathcal{M}$  by  $d(f, \mathcal{M}) := \inf\{\|f - h\|_{\mathcal{B}} : h \in \mathcal{M}\}$ . An element  $f_0 \in \mathcal{M}$  is said to be a best approximation to  $f$  from  $\mathcal{M}$  if  $\|f - f_0\|_{\mathcal{B}} = d(f, \mathcal{M})$ . A subset  $\mathcal{M}$  of  $\mathcal{B}$  is called a convex set if  $tf + (1-t)g \in \mathcal{M}$ , for all  $f, g \in \mathcal{M}$  and all  $t \in [0, 1]$ . It is easy to see that for any  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$ ,  $\mathcal{M}_{\mathbf{y}}$  defined by (2) is a closed convex subset and  $\mathcal{M}_0$  is a closed subspace of  $\mathcal{B}$ . In fact,  $\mathcal{M}_{\mathbf{y}}$  is a translate of  $\mathcal{M}_0$ :  $\mathcal{M}_{\mathbf{y}} = \mathcal{M}_0 + f$ , for each  $f \in \mathcal{M}_{\mathbf{y}}$ . This relation between  $\mathcal{M}_{\mathbf{y}}$  and  $\mathcal{M}_0$  allows us to develop a characterization of solutions of problem (3) in terms of best approximation from a subspace. To simplify our presentations, we state frequently used conditions in the following assumption.

(A1)  $\mathcal{B}$  is a Banach space with the dual space  $\mathcal{B}^*$  and  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , are linearly independent.

**Proposition 2** *Suppose that Assumption (A1) holds and  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$ . Then  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and 0 is a best approximation to  $\hat{f}$  from  $\mathcal{M}_0$ .*



Proposition 2 enables us to characterize a solution of problem (3) via identifying a best approximation from  $\mathcal{M}_0$  with continuous linear functionals by using the duality approach (the Hahn-Banach theorem), see, Braess (1986). To this end, we recall the definition of annihilators of subsets in Banach spaces. Let  $\mathcal{M}$  and  $\mathcal{M}'$  be subsets of  $\mathcal{B}$  and  $\mathcal{B}^*$ , respectively. According to Megginson (1998), the annihilator of  $\mathcal{M}$  in  $\mathcal{B}^*$  is defined by  $\mathcal{M}^\perp := \{\nu \in \mathcal{B}^* : \langle \nu, f \rangle_{\mathcal{B}} = 0, \text{ for all } f \in \mathcal{M}\}$ . The annihilator of  $\mathcal{M}'$  in  $\mathcal{B}$  is defined by  ${}^\perp\mathcal{M}' := \{f \in \mathcal{B} : \langle \nu, f \rangle_{\mathcal{B}} = 0, \text{ for all } \nu \in \mathcal{M}'\}$ . Proposition 2 together with Corollary 2.3 in Braess (1986) leads to the next proposition.

**Proposition 3** *Suppose that Assumption (A1) holds and  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$ . Then  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there is a continuous linear functional  $\nu \in \mathcal{M}_0^\perp$  such that*

$$\|\nu\|_{\mathcal{B}^*} = 1 \quad \text{and} \quad \langle \nu, \hat{f} \rangle_{\mathcal{B}} = \|\hat{f}\|_{\mathcal{B}}. \quad (6)$$

We next identify the subspace  $\mathcal{M}_0^\perp$  of  $\mathcal{B}^*$  with the linear span of the finite number of continuous linear functionals  $\mathcal{V}_m := \{\nu_j : j \in \mathbb{N}_m\}$ . The following lemma may be proved by Proposition 2.6.6 of Megginson (1998).

**Lemma 4** *If Assumption (A1) holds, then  $\mathcal{M}_0^\perp = \text{span } \mathcal{V}_m$ , where  $\mathcal{V}_m$  is defined above.*

Lemma 4 indicates that subspace  $\mathcal{M}_0^\perp$  is of finite dimension. This is a consequence of the fact that the number of continuous linear functionals appearing in  $\mathcal{M}_0$  is finite. Lemma 4 together with Proposition 3 leads to a solution representation of problem (3).

**Proposition 5** *Suppose that Assumption (A1) holds and  $\mathbf{y} \in \mathbb{R}^m$ . Then  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that the linear functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  satisfies equations (6).*

We now turn to establishing the representer theorem for a solution of problem (3) by using a convex analytic approach. Specifically, as a special convex programming problem with constraints, problem (3) can be solved by the Lagrange multiplier method. We recall the notion of the subdifferential of a convex function in a Banach space  $\mathcal{B}$ . For a convex function  $\phi : \mathcal{B} \rightarrow \mathbb{R} \cup \{+\infty\}$ , the subdifferential  $\partial\phi(f)$  of  $\phi$  at  $f$  is defined by

$$\partial\phi(f) := \{\nu \in \mathcal{B}^* : \phi(g) - \phi(f) \geq \langle \nu, g - f \rangle_{\mathcal{B}}, \text{ for all } g \in \mathcal{B}\}. \quad (7)$$

In convex programming, the Lagrange multiplier method provides a necessary and sufficient condition for solutions of optimization problems with constraints (Zălinescu, 2002). This leads to an alternative form of the representer theorem for a solution of problem (3).

**Theorem 6** *Suppose that Assumption (A1) holds and  $\mathbf{y} \in \mathbb{R}^m$ . Then  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\sum_{j \in \mathbb{N}_m} c_j \nu_j \in \partial\|\cdot\|_{\mathcal{B}}(\hat{f})$ .*

Both Proposition 5 and Theorem 6 reveal that the MNI problem with a finite number of data points, as a minimization problem in an infinite dimensional space, can be transferred to one in a finite dimensional space about finitely many coefficients  $c_j$ ,  $j \in \mathbb{N}_m$ . Although the two representer theorems are derived from two different viewpoints and described by different mathematical tools, they are intimately connected to each other. The bridge of these two viewpoints is the known result (Cioranescu, 1990) that the subdifferential of the norm  $\|\cdot\|_{\mathcal{B}}$  at  $\hat{f}$  coincides with the set of all the functionals satisfying equations (6). That is, for each  $f \in \mathcal{B} \setminus \{0\}$ , there holds

$$\partial\|\cdot\|_{\mathcal{B}}(f) = \{\nu \in \mathcal{B}^* : \|\nu\|_{\mathcal{B}^*} = 1, \langle \nu, f \rangle_{\mathcal{B}} = \|f\|_{\mathcal{B}}\}. \quad (8)$$

Another well-known notion related to the functionals satisfying equations (6) is the peak functional. For a real Banach space  $\mathcal{B}$  and for each  $\nu \in \mathcal{B}^*$ , there holds  $\langle \nu, f \rangle_{\mathcal{B}} \leq \|\nu\|_{\mathcal{B}^*} \|f\|_{\mathcal{B}}$ , for all  $f \in \mathcal{B}$ . For a fixed  $f \in \mathcal{B}$ , we are particularly interested in identifying functionals  $\nu \in \mathcal{B}^* \setminus \{0\}$  that allow  $\langle \nu, f \rangle_{\mathcal{B}}$  assuming the upper bound in the above inequality. We say that a functional  $\nu \in \mathcal{B}^*$  peaks at  $f \in \mathcal{B}$  if  $\langle \nu, f \rangle_{\mathcal{B}} = \|\nu\|_{\mathcal{B}^*} \|f\|_{\mathcal{B}}$ . This gives rise to the notion of the duality mapping on a Banach space (Cioranescu, 1990). Specifically, the duality mapping  $\mathcal{J}$  from  $\mathcal{B}$  to the collection of all subsets in  $\mathcal{B}^*$  is defined for all  $f \in \mathcal{B}$  by  $\mathcal{J}(f) := \{\nu \in \mathcal{B}^* : \|\nu\|_{\mathcal{B}^*} = \|f\|_{\mathcal{B}}, \langle \nu, f \rangle_{\mathcal{B}} = \|\nu\|_{\mathcal{B}^*} \|f\|_{\mathcal{B}}\}$ . A solution of problem (3) may be described by these functionals. In the next proposition, we summarize various solution representations of problem (3), developed by using various notions. This proposition may be proved by using the relations among these notions.

**Proposition 7** *Suppose that Assumption (A1) holds. Let  $\mathbf{y} \in \mathbb{R}^m$  and  $\hat{f} \in \mathcal{B}$ . Then the following statements are equivalent:*

- (i)  $\hat{f}$  is a solution of problem (3) with  $\mathbf{y}$ .
- (ii)  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that the linear functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  satisfies  $\|\nu\|_{\mathcal{B}^*} = 1$  and  $\langle \nu, \hat{f} \rangle_{\mathcal{B}} = \|\hat{f}\|_{\mathcal{B}}$ .
- (iii)  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that the linear functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  peaks at  $\hat{f}$ , that is,  $\langle \nu, \hat{f} \rangle_{\mathcal{B}} = \|\nu\|_{\mathcal{B}^*} \|\hat{f}\|_{\mathcal{B}}$ .
- (iv)  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\sum_{j \in \mathbb{N}_m} c_j \nu_j \in \mathcal{J}(\hat{f})$ .
- (v)  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\sum_{j \in \mathbb{N}_m} c_j \nu_j \in \partial\|\cdot\|_{\mathcal{B}}(\hat{f})$ .

Proposition 7 gives four characterizations of a solution of problem (3), which will serve as a basis for further developing the representer theorems.

We next consider the special case when  $\mathcal{B}$  is a smooth Banach space. In this case, the representer theorem can enjoy a nice simple form. To this end, we recall that the norm  $\|\cdot\|_{\mathcal{B}}$  is said to be Gâteaux differentiable at  $f \in \mathcal{B} \setminus \{0\}$  if for all  $h \in \mathcal{B}$ , the limits  $\lim_{t \rightarrow 0} \frac{\|f+th\|_{\mathcal{B}} - \|f\|_{\mathcal{B}}}{t}$  exist. If the norm  $\|\cdot\|_{\mathcal{B}}$  is Gâteaux differentiable at  $f \in \mathcal{B} \setminus \{0\}$ , then there exists a continuous linear functional, denoted by  $\mathcal{G}(f)$ , in  $\mathcal{B}^*$  such that

$$\langle \mathcal{G}(f), h \rangle_{\mathcal{B}} = \lim_{t \rightarrow 0} \frac{\|f+th\|_{\mathcal{B}} - \|f\|_{\mathcal{B}}}{t}, \quad \text{for all } h \in \mathcal{B}. \quad (9)$$

We call  $\mathcal{G}(f)$  the Gâteaux derivative of  $\|\cdot\|_{\mathcal{B}}$  at  $f \in \mathcal{B}$ . It follows from (9) that  $|\langle \mathcal{G}(f), h \rangle_{\mathcal{B}}| \leq \|h\|_{\mathcal{B}}$ , for all  $h \in \mathcal{B}$ , and  $\langle \mathcal{G}(f), f \rangle_{\mathcal{B}} = \|f\|_{\mathcal{B}}$ , which leads to  $\|\mathcal{G}(f)\|_{\mathcal{B}^*} = 1$ . Since for  $f = 0$ ,

the limit defined as in (9) does not exist, the Gâteaux derivative of the norm  $\|\cdot\|_{\mathcal{B}}$  can not be defined at  $f = 0$ . To simplify the presentation, we define the Gâteaux derivative of the norm  $\|\cdot\|_{\mathcal{B}}$  at  $f = 0$  by  $\mathcal{G}(f) := 0$ . A Banach space  $\mathcal{B}$  is said to be smooth if the norm  $\|\cdot\|_{\mathcal{B}}$  is Gâteaux differentiable at every  $f \in \mathcal{B} \setminus \{0\}$ . By employing the equivalence of (i) and (v) in Proposition 7 and showing that the subdifferential of  $\|\cdot\|_{\mathcal{B}}$  at any  $f \in \mathcal{B} \setminus \{0\}$  is the singleton  $\mathcal{G}(f)$ , we have the following special result.

**Theorem 8** *Suppose that Assumption (A1) holds and  $\mathbf{y} \in \mathbb{R}^m$ . If  $\mathcal{B}$  is smooth, then  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\mathcal{G}(\hat{f}) = \sum_{j \in \mathbb{N}_m} c_j \nu_j$ .*

We next derive explicit representer theorems for problem (3). Theorems 6 and 8 provide implicit representer theorems for a solution  $\hat{f}$  of problem (3). It would be more informative to have an explicit representation. To this end, we establish a duality lemma.

**Lemma 9** *Suppose that  $\mathcal{B}$  is a Banach space with the dual space  $\mathcal{B}^*$ ,  $f \in \mathcal{B} \setminus \{0\}$  and  $\nu \in \mathcal{B}^* \setminus \{0\}$ . Then*

$$\frac{\nu}{\|\nu\|_{\mathcal{B}^*}} \in \partial\|\cdot\|_{\mathcal{B}}(f) \quad \text{if and only if} \quad \frac{f}{\|f\|_{\mathcal{B}}} \in \partial\|\cdot\|_{\mathcal{B}^*}(\nu). \quad (10)$$

Lemma 9 enables us to “solve”  $\hat{f}$  from the implicit representer theorems. We present these explicit representer theorems next.

**Theorem 10** *Suppose that Assumption (A1) holds and  $\mathbf{y} \in \mathbb{R}^m$ . Then  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that*

$$\hat{f} \in \gamma \partial\|\cdot\|_{\mathcal{B}^*} \left( \sum_{j \in \mathbb{N}_m} c_j \nu_j \right), \quad \text{with } \gamma := \left\| \sum_{j \in \mathbb{N}_m} c_j \nu_j \right\|_{\mathcal{B}^*}. \quad (11)$$

We may get a special representer theorem when  $\mathcal{B}$  is a Banach space having the *smooth* dual space  $\mathcal{B}^*$ . It can be proved by employing Theorem 10 and noticing that the subdifferential of  $\|\cdot\|_{\mathcal{B}^*}$  at nonzero  $\nu \in \mathcal{B}^*$  is the singleton  $\mathcal{G}^*(\nu)$ , the Gâteaux derivative of the norm  $\|\cdot\|_{\mathcal{B}^*}$  at  $\nu$ .

**Remark 11** *If the dual space  $\mathcal{B}^*$  of  $\mathcal{B}$  is smooth, then  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that*

$$\hat{f} = \gamma \mathcal{G}^* \left( \sum_{j \in \mathbb{N}_m} c_j \nu_j \right). \quad (12)$$

Next, we establish the representer theorems in a special case when the Banach space  $\mathcal{B}$  has a pre-dual space  $\mathcal{B}_*$  and  $\nu_j \in \mathcal{B}_*$ ,  $j \in \mathbb{N}_m$ . We state the following assumption.

(A2)  $\mathcal{B}$  is a Banach space having a pre-dual space  $\mathcal{B}_*$  and  $\nu_j \in \mathcal{B}_*$ ,  $j \in \mathbb{N}_m$ , are linearly independent.

**Theorem 12** *Suppose that Assumption (A2) holds and  $\mathbf{y} \in \mathbb{R}^m$ . Then  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\hat{f} \in \rho \partial \|\cdot\|_{\mathcal{B}_*} \left( \sum_{j \in \mathbb{N}_m} c_j \nu_j \right)$ , with  $\rho := \left\| \sum_{j \in \mathbb{N}_m} c_j \nu_j \right\|_{\mathcal{B}_*}$ .*

Theorem 12 may be reduced to a nice simple form when  $\mathcal{B}$  is a Banach space having a smooth pre-dual space  $\mathcal{B}_*$ . We denote by  $\mathcal{G}_*(\nu)$  the Gâteaux derivative of the norm  $\|\cdot\|_{\mathcal{B}_*}$  at  $\nu \in \mathcal{B}_*$ .

**Remark 13** *If  $\mathcal{B}$  has a smooth pre-dual space  $\mathcal{B}_*$  and  $\mathbf{y} \in \mathbb{R}^m$ , then  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\hat{f} = \rho \mathcal{G}_* \left( \sum_{j \in \mathbb{N}_m} c_j \nu_j \right)$ .*

In the remaining part of this section, we consider several specific cases of practical interest. An immediate consequence of Remark 11 is the classical representer theorem for MNI in an RKHS (Wendland, 2005). In this special case, the functionals  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , are the point-evaluation functionals  $\delta_{x_j}$ ,  $j \in \mathbb{N}_m$ , where  $x_j$ ,  $j \in \mathbb{N}_m$ , are points in an input set  $X$ . We call a Hilbert space  $\mathcal{H}$  of functions on  $X$  an RKHS if the point-evaluation functionals are continuous on  $\mathcal{H}$ . According to the Riesz representation theorem, for each RKHS  $\mathcal{H}$  there exists a unique reproducing kernel  $K : X \times X \rightarrow \mathbb{R}$  satisfying  $K(x, \cdot) \in \mathcal{H}$  for all  $x \in X$  and the reproducing property  $f(x) = \langle f, K(x, \cdot) \rangle_{\mathcal{H}}$ , for all  $x \in X$  and all  $f \in \mathcal{H}$ . By the reproducing property, the kernel  $K$  provides a closed-form function representation for the point-evaluation functionals, that is,  $\nu_j := K(x_j, \cdot)$ ,  $j \in \mathbb{N}_m$ . Remark 11 ensures that the unique solution  $\hat{f}$  of problem (3) with  $\mathbf{y}$  can be represented by  $\hat{f} = \left\| \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot) \right\|_{\mathcal{H}} \mathcal{G} \left( \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot) \right)$  for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ . It is known that when  $\mathcal{H}$  is a Hilbert space,  $f = \|f\|_{\mathcal{H}} \mathcal{G}(f)$  for all  $f \in \mathcal{H}$ . By using this equation with  $f := \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot)$ , we get a special form  $\hat{f} = \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot)$  when  $\mathcal{H}$  is a RKHS with the kernel  $K$ .

We next consider the MNI problem in a Banach space  $\mathcal{B}$  that is uniformly Fréchet smooth and uniformly convex. Such a space can provide a semi-inner-product as a useful tool for representing the solution of the problem. The norm of a normed space  $\mathcal{B}$  is said to be uniformly Fréchet differentiable if the limit in (9) exists for every  $f \in \mathcal{B} \setminus \{0\}$  and  $h \in \mathcal{B}$ , and the convergence is uniform for all  $f, h$  in the unit sphere of  $\mathcal{B}$ . Accordingly, a normed space is uniformly Fréchet smooth if its norm is uniformly Fréchet differentiable. A normed space  $\mathcal{B}$  is uniformly convex if for all  $\varepsilon > 0$  there exists a  $\delta > 0$  such that  $\|f + g\|_{\mathcal{B}} \leq 2 - \delta$  for all  $f, g$  in the unit sphere of  $\mathcal{B}$  with  $\|f - g\|_{\mathcal{B}} \geq \varepsilon$ . The Milman-Pettis Theorem (Megginson, 1998) states that every uniformly convex Banach space  $\mathcal{B}$  is reflexive.

It follows from Giles (1967) that for a smooth Banach space  $\mathcal{B}$ , there exists a unique semi-inner-product  $[\cdot, \cdot]_{\mathcal{B}} : \mathcal{B} \times \mathcal{B} \rightarrow \mathbb{R}$  that induces its norm by  $\|\cdot\|_{\mathcal{B}} := [\cdot, \cdot]_{\mathcal{B}}^{1/2}$ . Note that the semi-inner-product  $[\cdot, \cdot]_{\mathcal{B}}$  is *not* linear with respect to its second variable. For each  $g \in \mathcal{B}$ , we introduce the linear functional  $\nu$  on  $\mathcal{B}$  by  $\nu(f) := [f, g]_{\mathcal{B}}$ , for all  $f \in \mathcal{B}$ . Then by the Cauchy-Schwarz inequality, the linear functional  $\nu$  is continuous. Following Zhang et al. (2009), this functional, denoted by  $g^{\sharp}$ , is called the dual element of  $g$ . That is,  $\langle g^{\sharp}, f \rangle_{\mathcal{B}} = [f, g]_{\mathcal{B}}$  for all  $f \in \mathcal{B}$ . A generalization of the Riesz Representation Theorem in Banach spaces given in Giles (1967) states that if  $\mathcal{B}$  is a uniformly Fréchet smooth and uniformly convex Banach space, then for each  $\nu \in \mathcal{B}^*$ , there exists a unique  $g \in \mathcal{B}$  such

that  $\nu = g^\sharp$  and  $\|\nu\|_{\mathcal{B}^*} = \|g\|_{\mathcal{B}}$ . Accordingly, the mapping  $f \rightarrow f^\sharp$  is bijective from  $\mathcal{B}$  to  $\mathcal{B}^*$ . There is a well-known relation between uniform Fréchet smoothness and uniform convexity (Megginson, 1998). Specifically, a normed space is uniformly convex if and only if its dual space is uniformly Fréchet smooth, and is uniformly Fréchet smooth if and only if its dual space is uniformly convex. Hence, if a Banach space  $\mathcal{B}$  is uniformly Fréchet smooth and uniformly convex, then so is its dual space  $\mathcal{B}^*$ .

As a consequence of Theorem 8 and Remark 11, we get the following representer theorems for the MNI problem in a uniformly Fréchet smooth and uniformly convex Banach space  $\mathcal{B}$ . In this case, the linearly independent functionals  $\nu_j$  can be identified with  $g_j^\sharp$  for  $g_j \in \mathcal{B}$ ,  $j \in \mathbb{N}_m$ . We state the following assumption.

(A3)  $\mathcal{B}$  is a uniformly Fréchet smooth and uniformly convex Banach space with the dual space  $\mathcal{B}^*$  and  $g_j \in \mathcal{B}$ ,  $j \in \mathbb{N}_m$ .

**Theorem 14** *Suppose that Assumption (A3) holds and  $g_j^\sharp \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , are linearly independent. Then*

(1) *problem (3) with  $\mathbf{y} \in \mathbb{R}^m$  has a unique solution  $\hat{f}$  such that  $\hat{f}^\sharp = \sum_{j \in \mathbb{N}_m} c_j g_j^\sharp$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ ;*

(2) *problem (3) with  $\mathbf{y} \in \mathbb{R}^m$  has a unique solution  $\hat{f}$  in the form  $\hat{f} = \left( \sum_{j \in \mathbb{N}_m} c_j g_j^\sharp \right)^\sharp$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

The explicit representation stated in statement (2) of Theorem 14 can also be obtained directly from statement (1) with the following simple fact.

**Lemma 15** *If  $\mathcal{B}$  is a uniformly Fréchet smooth and uniformly convex Banach space, then for any  $f \in \mathcal{B}$ , there holds  $f^{\sharp\sharp} = f$ .*

As a special example, we consider the MNI problem in the Banach space  $\ell_p(\mathbb{N})$  with  $1 < p < +\infty$ , which is the Banach space of all real sequences  $\mathbf{x} := (x_j : j \in \mathbb{N})$ , with  $\|\mathbf{x}\|_p := (\sum_{j \in \mathbb{N}} |x_j|^p)^{1/p} < +\infty$ . The space  $\ell_p(\mathbb{N})$  is uniformly Fréchet smooth and uniformly convex and has  $\ell_q(\mathbb{N})$  as its dual space, where  $1/p + 1/q = 1$ . The dual bilinear form  $\langle \cdot, \cdot \rangle_{\ell_p}$  on  $\ell_q(\mathbb{N}) \times \ell_p(\mathbb{N})$  is defined by  $\langle \mathbf{u}, \mathbf{x} \rangle_{\ell_p} := \sum_{j \in \mathbb{N}} u_j x_j$ , for all  $\mathbf{u} := (u_j : j \in \mathbb{N}) \in \ell_q(\mathbb{N})$  and all  $\mathbf{x} := (x_j : j \in \mathbb{N}) \in \ell_p(\mathbb{N})$ . In this case, we suppose that  $\mathbf{u}_j$ ,  $j \in \mathbb{N}_m$ , are a finite number of linearly independent elements in  $\ell_q(\mathbb{N})$  and the operator  $\mathcal{L} : \ell_p(\mathbb{N}) \rightarrow \mathbb{R}^m$ , defined by (1), has the form

$$\mathcal{L}(\mathbf{x}) := [\langle \mathbf{u}_j, \mathbf{x} \rangle_{\ell_p} : j \in \mathbb{N}_m], \quad \text{for all } \mathbf{x} \in \ell_p(\mathbb{N}). \quad (13)$$

Applying statement (2) of Theorem 14 to the MNI problem with  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$  in  $\ell_p(\mathbb{N})$  and noticing that for any  $\mathbf{u} = (u_j : j \in \mathbb{N}) \in \ell_q(\mathbb{N})$ ,  $\mathbf{u}^\sharp = (u_j |u_j|^{q-2} / \|\mathbf{u}\|_q^{q-2} : j \in \mathbb{N})$ , we get the explicit representation of the unique solution  $\hat{\mathbf{x}} := (\hat{x}_j : j \in \mathbb{N})$  as  $\hat{x}_j = u_j |u_j|^{q-2} / \|\mathbf{u}\|_q^{q-2}$ , with  $\mathbf{u} = (u_j : j \in \mathbb{N}) := \sum_{j \in \mathbb{N}_m} c_j \mathbf{u}_j$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .

We next consider the MNI problem in two types of RKBSs  $\mathcal{B}$ . In these spaces, the functionals  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , also refer to point-evaluation functionals  $\delta_{x_j}$ ,  $j \in \mathbb{N}_m$ , where  $x_j$ ,  $j \in \mathbb{N}_m$ , are finite points in an input set  $X$ . The notion of RKBSs was originally introduced in Zhang et al. (2009) based on the semi-inner-product. We begin by reviewing the notion of semi-inner-product RKBSs. A Banach space  $\mathcal{B}$  of functions on a domain  $X$  is called a semi-inner-product RKBS if it is uniformly Fréchet smooth and uniformly

convex, and the point-evaluation functionals are continuous linear functionals on  $\mathcal{B}$ . If  $\mathcal{B}$  is a semi-inner-product RKBS, then there exists a unique (semi-inner-product) reproducing kernel  $G : X \times X \rightarrow \mathbb{R}$  satisfying  $G(x, \cdot) \in \mathcal{B}$  for all  $x \in X$  and the reproducing property  $f(x) = [f, G(x, \cdot)]_{\mathcal{B}}$ , for all  $x \in X$  and all  $f \in \mathcal{B}$ . Since a semi-inner-product RKBS is uniformly Fréchet smooth and uniformly convex, Theorems 14 lead to the following representer theorems for the MNI problem in a semi-inner-product RKBS. It suffices to note by the reproducing property that for each  $x \in X$ , the dual element  $G(x, \cdot)^{\sharp}$  of  $G(x, \cdot)$  coincides with the point-evaluation functional  $\delta_x$ .

**Corollary 16** *If  $\mathcal{B}$  is a semi-inner-product RKBS with the reproducing kernel  $G$  and  $x_j \in X$ ,  $j \in \mathbb{N}_m$ , such that  $G(x_j, \cdot)^{\sharp} \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , are linearly independent, then*

(1) *problem (3) with  $\mathbf{y} \in \mathbb{R}^m$  has a unique solution  $\hat{f}$  such that  $\hat{f}^{\sharp} = \sum_{j \in \mathbb{N}_m} c_j G(x_j, \cdot)^{\sharp}$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ ;*

(2) *problem (3) with  $\mathbf{y} \in \mathbb{R}^m$  has a unique solution  $\hat{f}$  such that  $\hat{f} = \left( \sum_{j \in \mathbb{N}_m} c_j G(x_j, \cdot)^{\sharp} \right)^{\sharp}$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

An alternative definition of RKBSs was introduced in Xu and Ye (2019). This definition is a natural generalization of RKHSs by replacing the inner product in the Hilbert spaces with the dual bilinear form in introducing the reproducing properties in RKBSs. We now apply Theorem 8 and Remark 11 to the MNI problem in such an RKBS, whose definition is given below. Suppose that  $\mathcal{B}$  is a Banach space of functions on  $X$  and the dual space  $\mathcal{B}^*$  is isometrically equivalent to a Banach space of functions on  $X'$ . A Banach space  $\mathcal{B}$  is called a right-sided RKBS and  $K : X \times X' \rightarrow \mathbb{R}$  is its right-sided reproducing kernel if  $K(x, \cdot) \in \mathcal{B}^*$  for all  $x \in X$  and  $f(x) = \langle K(x, \cdot), f \rangle_{\mathcal{B}}$ , for all  $x \in X$  and all  $f \in \mathcal{B}$ . In the framework of right-sided RKBSs, we consider the MNI problem with a finite number of point-evaluation functionals  $\delta_{x_j}$ ,  $j \in \mathbb{N}_m$ , where  $x_j \in X$ ,  $j \in \mathbb{N}_m$ . The representer theorem for this case can be obtained directly from Theorem 8. We state the following assumption.

(A4)  $\mathcal{B}$  is a right-sided RKBS with the right-sided reproducing kernel  $K$  and  $x_j \in X$ ,  $j \in \mathbb{N}_m$ .

**Corollary 17** *Suppose that Assumption (A4) holds and  $K(x_j, \cdot) \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , are linearly independent. If  $\mathcal{B}$  is reflexive, strictly convex, and smooth, then problem (3) with  $\mathbf{y} \in \mathbb{R}^m$  has a unique solution  $\hat{f}$  such that  $\mathcal{G}(\hat{f}) = \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

An explicit formula derived from Remark 11 is presented next.

**Theorem 18** *Suppose that Assumption (A4) holds and  $K(x_j, \cdot) \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , are linearly independent. If  $\mathcal{B}$  is reflexive and strictly convex, then problem (3) with  $\mathbf{y} \in \mathbb{R}^m$  has a unique solution  $\hat{f}$  in the form  $\hat{f} = \rho \mathcal{G}^* \left( \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot) \right)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , where  $\rho := \left\| \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot) \right\|_{\mathcal{B}^*}$ .*

Finally, we consider the MNI problem in the Banach space  $\ell_1(\mathbb{N})$ , which was studied in Cheng and Xu (2020). MNI in  $\ell_1(\mathbb{N})$  is a “compressed sensing” (Candès et al., 2006; Donoho, 2006) in the infinite dimensional space. The difference between these two problems is that the latter obtains a sparse solution for a finite dimensional problem while the former

secures a sparse solution in a finite dimensional space for an infinite dimensional problem. The use of the  $\ell_1$  norm in machine learning can induce sparsity in learning algorithms, see (Unser, 2016, 2019; Zhu et al., 2004). For example, the 1-norm support vector machine uses the  $\ell_1$  norm to promote sparsity of its solution for approximate accuracy and computational efficiency. Furthermore, RKBSs with the 1-norm were considered recently by Song et al. (2013) and Xu and Ye (2019) in developing sparse learning methods. It is known that  $\ell_1(\mathbb{N})$  is a typical RKBS and some RKBSs with the 1-norm are in fact isometrically isomorphic to  $\ell_1(\mathbb{N})$ . Note that  $\ell_1(\mathbb{N})$  has  $c_0$  as its pre-dual space (Benyamini and Lindenstrauss, 1972) and  $\ell_\infty(\mathbb{N})$  as its dual space. The space  $\ell_1(\mathbb{N})$  is not reflexive since the dual space of  $\ell_\infty(\mathbb{N})$  is not  $\ell_1(\mathbb{N})$ , and its pre-dual space  $c_0$  is not smooth. The dual bilinear form  $\langle \cdot, \cdot \rangle_{\ell_1}$  on  $\ell_\infty(\mathbb{N}) \times \ell_1(\mathbb{N})$  is defined by  $\langle \mathbf{u}, \mathbf{x} \rangle_{\ell_1} := \sum_{j \in \mathbb{N}} u_j x_j$ , for all  $\mathbf{u} := (u_j : j \in \mathbb{N}) \in \ell_\infty(\mathbb{N})$  and all  $\mathbf{x} := (x_j : j \in \mathbb{N}) \in \ell_1(\mathbb{N})$ . In this case, the functionals in the MNI problem belong to  $c_0$ . Specifically, we suppose that  $\mathbf{u}_j, j \in \mathbb{N}_m$ , are a finite number of linearly independent elements in  $c_0$ . The operator  $\mathcal{L} : \ell_1(\mathbb{N}) \rightarrow \mathbb{R}^m$ , defined by (1), has the form

$$\mathcal{L}(\mathbf{x}) := [\langle \mathbf{u}_j, \mathbf{x} \rangle_{\ell_1} : j \in \mathbb{N}_m], \quad \text{for all } \mathbf{x} \in \ell_1(\mathbb{N}). \quad (14)$$

To obtain the representer theorem for this case from Theorem 12, we need to compute explicitly the subdifferentials of the norm  $\|\cdot\|_\infty$  of  $c_0$ . Let  $X$  be a vector space and  $\mathcal{V}$  a subset. The convex hull of  $\mathcal{V}$ , denoted by  $\text{co}(\mathcal{V})$ , is the collection of all convex combinations of elements of  $\mathcal{V}$ , that is,

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

For each  $\mathbf{u} \in c_0$ , we let  $\mathbb{N}(\mathbf{u})$  denote the index set on which the sequence  $\mathbf{u}$  attains its norm  $\|\mathbf{u}\|_\infty$ , that is,  $\|\mathbf{u}\|_\infty = |u_j|, j \in \mathbb{N}(\mathbf{u})$  and  $\|\mathbf{u}\|_\infty > |u_j|, j \notin \mathbb{N}(\mathbf{u})$ . For each  $\mathbf{u} \in c_0$ , since  $u_j$  tends to zero as  $j \rightarrow +\infty$ , we have that  $\#\mathbb{N}(\mathbf{u}) < +\infty$ . To present the subdifferentials of the norm  $\|\cdot\|_\infty$  of  $c_0$  at any  $\mathbf{u} := (u_j : j \in \mathbb{N}) \in c_0$ , we introduce a subset of  $\ell_1(\mathbb{N})$  as

$$\mathcal{V}(\mathbf{u}) := \{\text{sign}(u_j) \mathbf{e}_j : j \in \mathbb{N}(\mathbf{u})\}, \quad (15)$$

where for each  $j \in \mathbb{N}$ ,  $\mathbf{e}_j$  denotes the vector in  $\ell_1(\mathbb{N})$  whose  $j$ th component is equal to 1 and all other components are zero.

The following lemma which was essentially proved in Cheng and Xu (2020) presents the subdifferential of the norm  $\|\cdot\|_\infty$  of  $c_0$  at any nonzero  $\mathbf{u} := (u_j : j \in \mathbb{N}) \in c_0$ .

**Lemma 19** *If  $\mathbf{u} := (u_j : j \in \mathbb{N})$  is a nonzero element in  $c_0$  and  $\mathcal{V}(\mathbf{u})$  is defined by (15), then  $\partial\|\cdot\|_\infty(\mathbf{u}) = \text{co}(\mathcal{V}(\mathbf{u}))$ .*

Combining Lemma 19 and Theorem 12, we get the following result.

**Theorem 20** *Suppose that  $\mathbf{u}_j, j \in \mathbb{N}_m$ , are a finite number of linearly independent elements in  $c_0$ . Let  $\mathbf{y} \in \mathbb{R}^m$  and  $\mathcal{L}$  and  $\mathcal{M}_\mathbf{y}$  be defined by (14) and (2), respectively. Then  $\hat{\mathbf{x}} \in \ell_1(\mathbb{N})$  is a solution of problem (3) in  $\ell_1(\mathbb{N})$  with  $\mathbf{y}$  if and only if  $\hat{\mathbf{x}} \in \mathcal{M}_\mathbf{y}$  and there exist  $c_j \in \mathbb{R}, j \in \mathbb{N}_m$ , such that  $\hat{\mathbf{x}} \in \|\mathbf{u}\|_\infty \text{co}(\mathcal{V}(\mathbf{u}))$ , with  $\mathbf{u} := \sum_{j \in \mathbb{N}_m} c_j \mathbf{u}_j$ .*

#### 4. Representer Theorem Based Solution Methods for MNI

The representer theorems presented in the last section for the MNI problem (3) give only forms of the solutions for the problem, *not* providing methods to determine the coefficients  $c_j$  involved in the solution representations. We develop in this section approaches to determine these coefficients, leading to solution methods for problem (3) when the Banach space  $\mathcal{B}$  has a pre-dual space  $\mathcal{B}_*$  and  $\nu_j \in \mathcal{B}_*$ , for  $j \in \mathbb{N}_m$ . We will consider both cases when the pre-dual space is smooth and non-smooth. Proofs of these results will be presented in Appendix D.

As a preparation, we express the adjoint operator  $\mathcal{L}^*$  of  $\mathcal{L}$  defined by (1). According to the continuity of the linear operator  $\mathcal{L}$  on  $\mathcal{B}$ , there exists a unique bounded linear operator  $\mathcal{L}^* : \mathbb{R}^m \rightarrow \mathcal{B}^*$ , called the adjoint operator of  $\mathcal{L}$ , such that  $\langle \mathcal{L}^*(\mathbf{c}), f \rangle_{\mathcal{B}} = \langle \mathbf{c}, \mathcal{L}(f) \rangle_{\mathbb{R}^m}$  for all  $f \in \mathcal{B}$  and all  $\mathbf{c} := [c_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$ . It follows from definition (1) of  $\mathcal{L}$  that  $\langle \mathcal{L}^*(\mathbf{c}), f \rangle_{\mathcal{B}} = \sum_{j \in \mathbb{N}_m} c_j \langle \nu_j, f \rangle_{\mathcal{B}} = \left\langle \sum_{j \in \mathbb{N}_m} c_j \nu_j, f \right\rangle_{\mathcal{B}}$ , which leads to  $\mathcal{L}^*(\mathbf{c}) = \sum_{j \in \mathbb{N}_m} c_j \nu_j$ .

We first provide the complete solution of problem (3) in a Banach space having a smooth pre-dual space  $\mathcal{B}_*$ . In this case, the solution of problem (3) with data  $\mathbf{y}$  can be obtained by employing Remark 13 with the coefficients  $c_j$  involved in it being chosen as a solution of a system, possibly nonlinear. In the following presentations, we always assume that  $\mathcal{L}$  is the operator defined by (1) and  $\mathcal{L}^*$  the associated adjoint operator.

**Theorem 21** *Suppose that Assumption (A2) holds. If  $\mathcal{B}_*$  is smooth and  $\mathbf{y} := [y_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$ , then*

$$\hat{f} := \|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} \mathcal{G}_*(\mathcal{L}^*(\mathbf{c})), \quad (16)$$

*is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\mathbf{c} \in \mathbb{R}^m$  is a solution of the system*

$$\langle \nu_k, \|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} \mathcal{G}_*(\mathcal{L}^*(\mathbf{c})) \rangle_{\mathcal{B}} = y_k, \quad k \in \mathbb{N}_m. \quad (17)$$

There are two interesting special cases. The first one concerns the MNI problem in a Hilbert space, that is,  $\mathcal{B} = \mathcal{H}$  is a Hilbert space. In this case,  $\mathcal{H}_* = \mathcal{H}$  and the linearly independent functionals  $\nu_j$  can be identified with  $g_j \in \mathcal{H}$ , for  $j \in \mathbb{N}_m$ . We then introduce the Gram matrix as  $\mathbf{G} := [\langle g_j, g_k \rangle_{\mathcal{H}} : j, k \in \mathbb{N}_m]$ . In this special case, Theorem 21 implies a known result in Wendland (2005). To see this, substituting

$$\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} \mathcal{G}_*(\mathcal{L}^*(\mathbf{c})) = \mathcal{L}^*(\mathbf{c}) \quad (18)$$

into (16), we represent  $\hat{f} = \sum_{j \in \mathbb{N}_m} c_j g_j$ . Again substituting (18) into (17), we get that  $\langle g_k, \mathcal{L}^*(\mathbf{c}) \rangle_{\mathcal{H}} = y_k$ ,  $k \in \mathbb{N}_m$ . By the representation of the adjoint operator  $\mathcal{L}^*$ , these equations can be rewritten in the form  $\mathbf{G}\mathbf{c} = \mathbf{y}$ . Note that according to the linear independence of  $g_j \in \mathcal{H}$ ,  $j \in \mathbb{N}_m$ , the Gram matrix  $\mathbf{G}$  is symmetric and positive definite. Therefore, the linear system  $\mathbf{G}\mathbf{c} = \mathbf{y}$  has a unique solution. Hence, Theorem 21 ensures that  $\hat{f} := \sum_{j \in \mathbb{N}_m} c_j g_j$  is the unique solution of problem (3) with  $\mathbf{y}$  if and only if  $\mathbf{c} := [c_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$  is the solution of the linear system of equations  $\mathbf{G}\mathbf{c} = \mathbf{y}$ .

We remark that in the case of Hilbert spaces, the *infinite* dimensional MNI problem is reduced to solving an equivalent *finite* dimensional linear system. The only infinite dimensional component in the linear system is computing the entries of the Gram matrix  $\mathbf{G}$  which requires calculating the inner products of  $g_k$  and  $g_j$ , elements in the infinite dimensional space  $\mathcal{H}$ .



As shown above, the MNI problem in a Hilbert space is reduced to solving a *linear system*. However, in a Banach space, not Hilbert, the problem cannot be reduced to a linear system. This will be demonstrated in the next case, where  $\mathcal{B}$  is a uniformly Fréchet smooth and uniformly convex Banach space (thus, in this case  $\mathcal{B}_* = \mathcal{B}^*$ ).

**Corollary 22** *Suppose that Assumption (A3) holds and  $g_j^\sharp \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , are linearly independent. Let  $\mathbf{y} := [y_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$ . Then  $\hat{f} = (\mathcal{L}^*(\mathbf{c}))^\sharp$ , is the unique solution of problem (3) with  $\mathbf{y}$  if and only if  $\mathbf{c} \in \mathbb{R}^m$  is the solution of the system of equations*

$$[g_k^\sharp, \mathcal{L}^*(\mathbf{c})]_{\mathcal{B}^*} = y_k, \quad k \in \mathbb{N}_m. \quad (19)$$

Observing from above results, a solution of the MNI problem in a Banach space having a smooth pre-dual can be represented by a finite number of functionals  $\nu_j$ , whose coefficients can be obtained by solving a system of equations. The resulting systems of equations are generally *nonlinear* unless  $\mathcal{B}$  is a Hilbert space. In particular, in the Banach space  $\mathcal{B}$  defined by the semi-inner-product, equations (19) are truly nonlinear due to the nonlinearity of the semi-inner-product with respect to the second variable. Similar to the Hilbert space case, the infinite dimensional component of this case lies in the computation of the semi-inner-product of two elements of the infinite dimensional Banach space.

We next consider solving problem (3) in a Banach space having a *non-smooth* pre-dual space by making use the representer theorem obtained in section 3. In this case, we do not assume that the pre-dual space is smooth. The solution methods presented here is mainly a continuation of Theorem 12. Recall that Theorem 12 provides a characterization of a solution of problem (3) in the case when a Banach space  $\mathcal{B}$  has a non-smooth pre-dual space  $\mathcal{B}_*$ . However, the theorem does not furnish a way to obtain the  $m$  coefficients  $c_j$  involved in the solution representation. Our task is to show that the coefficients  $c_j$  can, in deed, be obtained by solving an optimization problem in  $\mathbb{R}^m$ . To this end, we introduce the finite dimensional minimization problem with  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$  as

$$\inf \{ \|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} : \langle \mathbf{c}, \mathbf{y} \rangle_{\mathbb{R}^m} = 1, \mathbf{c} \in \mathbb{R}^m \}. \quad (20)$$

Note that minimization problem (20) is a somewhat twisted version of the compressed sensing problem (Candès et al., 2006; Donoho, 2006).

We begin with characterizing the solutions of (20) by standard arguments in convex analysis. The solutions of the minimization problem (20) can be first characterized by the Lagrange multiplier method, stated in Lemma 69, and the chain rule of the subdifferential (Showalter, 1997).

**Proposition 23** *Suppose that Assumption (A2) holds and  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$ . Then  $\hat{\mathbf{c}} \in \mathbb{R}^m$  is a solution of the minimization problem (20) with  $\mathbf{y}$  if and only if  $\langle \hat{\mathbf{c}}, \mathbf{y} \rangle_{\mathbb{R}^m} = 1$ , and there exist  $\lambda \in \mathbb{R}$  and  $f \in \partial \|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*(\hat{\mathbf{c}}))$  such that  $\mathcal{L}(f) = \lambda \mathbf{y}$ .*

We next present an alternative characterization of solutions of problem (20), which will be used to reveal the relation of solutions of minimization problems (3) and (20).

**Proposition 24** *Suppose that Assumption (A2) holds and  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$ . Then  $\hat{\mathbf{c}} \in \mathbb{R}^m$  is a solution of the minimization problem (20) with  $\mathbf{y}$  if and only if*

$$\|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}^{-1} \partial \|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*(\hat{\mathbf{c}})) \cap \mathcal{M}_{\mathbf{y}} \neq \emptyset. \quad (21)$$

We show below that the coefficients  $c_j$  appearing in the representer theorem (Theorem 12) can be obtained by solving the finite dimensional minimization problem (20). This yields a complete solution of problem (3) in a Banach space having a (non-smooth) pre-dual space.

**Theorem 25** *Suppose that Assumption (A2) holds and  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$ . Then  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}^{-1} \partial \|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*(\hat{\mathbf{c}})) \cap \mathcal{M}_{\mathbf{y}}$  for a solution  $\hat{\mathbf{c}}$  of problem (20) with  $\mathbf{y}$ .*

Theorem 25 provides a road map for finding a solution of problem (3). We describe major steps for solving the problem as follows:

Step 1: Find the solution  $\mathbf{c} := [c_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$  of the optimization problem (20).

Step 2: Construct  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  using components  $c_j$  of  $\mathbf{c}$  obtained from step 1.

Step 3: Find an element  $g \in \partial \|\cdot\|_{\mathcal{B}_*}(\nu)$  which satisfies  $\mathcal{L}(g) = \|\nu\|_{\mathcal{B}_*} \mathbf{y}$ , with  $\nu$  constructed in step 2.

Step 4: Obtain a solution of problem (3) by  $\hat{f} := \frac{g}{\|\nu\|_{\mathcal{B}_*}}$ , using  $\nu$  and  $g$  obtained respectively from steps 2 and 3.

Actual implementation of the above procedure requires further investigation. Note that although the minimization problem (20) in step 1 is of finite dimension, it still involves computation of the norm  $\|\cdot\|_{\mathcal{B}_*}$ , a hidden infinite dimensional component. Moreover, step 3 also involves solving an infinite dimensional problem. In order to make the above scheme implementable, we have to deal with these hidden infinite dimensional components. One could use approximation approaches to replace the infinite dimensional components by finite dimensional ones. This approach will introduce “truncation errors”, which we do *not* adopt here. Our idea is to make use intrinsic properties of these infinite dimensional components to remove their berries, developing *equivalent* implementable finite dimensional schemes.

Our approach to be described in section 5 is inspired by a recent result presented in Cheng and Xu (2020), where problem (3) with  $\mathcal{B} = \ell_1(\mathbb{N})$  was solved by a different approach. In this special case, the infinite dimensional components we mentioned above can be removed. This benefits from the characterization of the space  $c_0$ , a pre-dual space of  $\ell_1(\mathbb{N})$ . Firstly, the minimization problem (20) was reformulated as a linear programming problem. Specifically, suppose that  $\mathbf{u}_j$ ,  $j \in \mathbb{N}_m$ , are  $m$  given linearly independent elements in  $c_0$  and the operator  $\mathcal{L} : \ell_1(\mathbb{N}) \rightarrow \mathbb{R}^m$  is defined by (14). Instead of solving the minimization problem (20), it was proposed to solve an equivalent dual problem  $\sup\{\langle \mathbf{c}, \mathbf{y} \rangle_{\mathbb{R}^m} / \|\sum_{j \in \mathbb{N}_m} c_j \mathbf{u}_j\|_{\infty} : \mathbf{c} = [c_j : j \in \mathbb{N}_m] \in \mathbb{R}^m\}$ . It was proved there that the unit sphere in  $\mathbb{R}^m$  under the norm  $\|\cdot\|_*$ , defined by  $\|\mathbf{c}\|_* := \|\sum_{j \in \mathbb{N}_m} c_j \mathbf{u}_j\|_{\infty}$ ,  $\mathbf{c} \in \mathbb{R}^m$ , is the surface of a convex polytope, which are formed by a finite number of planes. Hence, the dual optimization problem is equivalent to a linear programming problem: finding a maximizer of the linear function  $g(\mathbf{c}) := \langle \mathbf{c}, \mathbf{y} \rangle_{\mathbb{R}^m}$ ,  $\mathbf{c} \in \mathbb{R}^m$  on the unit sphere  $\{\mathbf{c} : \mathbf{c} \in \mathbb{R}^m, \|\mathbf{c}\|_* = 1\}$ . Moreover, Lemma 19 shows that finding the subdifferentials of the norm  $\|\cdot\|_{\infty}$  of  $c_0$  at a nonzero element  $\mathbf{u} \in c_0$  is of finite dimension and any vector in the subdifferentials has at most finite many nonzero components. Therefore, a solution of the MNI problem in  $\ell_1(\mathbb{N})$  can be obtained by solving a linear system of  $m$  coefficients.

Although the MNI problem in  $\ell_1(\mathbb{N})$  can be solved as a truly finite dimensional problem, as described above, solving the resulting linear programming problem requires an exponential order  $\mathcal{O}(2^m)$  of computational costs, where  $m$  is the number of interpolation conditions

involved in the problem. When  $m$  is large, which is often the case in data science, this method is not computationally efficient. It is desirable to develop alternative representations of solutions of the MNI problem convenient for algorithmic development. Motivated by the success of the fixed-point approach used in machine learning (Li et al., 2020; Li et al., 2018, 2019; Polson et al., 2015), image processing (Chen et al., 2013; Li et al., 2012; Li et al., 2015; Lu et al., 2016; Micchelli et al., 2011), medical imaging (Krol et al., 2012; Li et al., 2015; Zheng et al., 2019) and solutions of inverse problems (Fan et al., 2014; Jin and Lu, 2014), we will develop representations of a solution of the MNI problem or the regularization problem in a Banach space, as fixed-points of nonlinear maps defined by proximity operators of functions involved in the problem. The fixed-point formulation well fits for designing iterative algorithms. Difficulty of developing implementable algorithms for this problem in a Banach space lies in infinite dimensional components of the problem. This challenge motivates us to develop a finite dimensional fixed-point approach to solve the MNI problem in the special Banach space  $\ell_1(\mathbb{N})$  by making use of special structures of this space and its pre-dual space. We present this approach in section 5. Extension of this approach to a general non-smooth Banach space will be a future research topic.

To close this section, we present the infimum of the MNI problem in a Banach space as a result of the explicit representer theorems. The next theorem identifies the infimum with the norm of the functional appearing in the explicit solution representation.

**Theorem 26** *Suppose that Assumption (A1) holds and  $\mathbf{y} \in \mathbb{R}^m$ . If  $\hat{f}$  is a solution of problem (3) with  $\mathbf{y}$ , which has either the form (11) or (12) with the coefficients  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , then  $\|\hat{f}\|_{\mathcal{B}} = \|\sum_{j \in \mathbb{N}_m} c_j \nu_j\|_{\mathcal{B}^*}$ .*

When the Banach space  $\mathcal{B}$  has a pre-dual space  $\mathcal{B}_*$  and  $\nu_j \in \mathcal{B}_*$ , for  $j \in \mathbb{N}_m$ , approaches were developed above to determine the coefficients  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , appearing in the solution representations. Accordingly, the infimum of problem (3) can be obtained from the resulting coefficients. The following result may be proved by Theorem 21 and similar arguments to those used in the proof of Theorem 26. We omit details of the proof.

**Theorem 27** *If Assumption (A2) holds,  $\mathcal{B}_*$  is smooth and  $\mathbf{y} \in \mathbb{R}^m$ , then the infimum  $m_0$  of problem (3) with  $\mathbf{y}$  has the form  $m_0 = \|\sum_{j \in \mathbb{N}_m} c_j \nu_j\|_{\mathcal{B}_*}$ , where  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , are the solution of the system (17) of equations.*

For the case that the pre-dual space  $\mathcal{B}_*$  of the Banach space  $\mathcal{B}$  may not be smooth, we have the following representation of the infimum by employing Theorem 25 and arguments similar to those used in the proof of Theorem 26. We also omit details of the proof.

**Theorem 28** *If Assumption (A2) holds and  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$ , then the infimum  $m_0$  of problem (3) with  $\mathbf{y}$  has the form  $m_0 = \|\sum_{j \in \mathbb{N}_m} \hat{c}_j \nu_j\|_{\mathcal{B}_*}^{-1}$ , where  $\hat{\mathbf{c}} := [\hat{c}_j : j \in \mathbb{N}_m]$  is a solution of the minimization problem (20) with  $\mathbf{y}$ .*

For the special cases discussed in section 3, infimum results similar to those stated in the above theorems remain valid. That is, in all the cases considered there, the infimum of the MNI problem is equal to the norm of the functional appearing in each corresponding explicit solution representation. We state these results below.

**Remark 29** *If  $\mathcal{B}$  is a uniformly Fréchet smooth and uniformly convex Banach space and  $\mathcal{B}^*$  is its dual space, then the infimum  $m_0$  of problem (3) with  $\mathbf{y}$  has the form  $m_0 = \|\sum_{j \in \mathbb{N}_m} c_j g_j^\sharp\|_{\mathcal{B}^*}$ , where  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , are the solution of the system (19) of equations.*

**Remark 30** *If  $\mathcal{B}$  is a semi-inner-product RKBS with the semi-inner-product reproducing kernel  $G$  and  $\mathcal{B}^*$  is its dual space, then the infimum  $m_0$  of problem (3) with  $\mathbf{y}$  has the form  $m_0 = \|\sum_{j \in \mathbb{N}_m} c_j G(x_j, \cdot)^\sharp\|_{\mathcal{B}^*}$ , where  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , are the solution of the system (19) of equations with  $g_k := G(x_k, \cdot)$ ,  $k \in \mathbb{N}_m$ .*

**Remark 31** *Suppose that  $\mathcal{B}$  is a right-sided RKBS with the right-sided reproducing kernel  $K$  and  $\mathcal{B}^*$  is its dual space. If  $\mathcal{B}$  is reflexive and strictly convex, then the infimum  $m_0$  of problem (3) with  $\mathbf{y}$  has the form  $m_0 = \|\sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot)\|_{\mathcal{B}^*}$ , where  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , are the solution of the system (17) of equations with  $\nu_j := K(x_j, \cdot)$ ,  $j \in \mathbb{N}_m$ .*

**Remark 32** *The infimum  $m_0$  of problem (3) in  $\ell_1(\mathbb{N})$  has the form  $m_0 = \|\sum_{j \in \mathbb{N}_m} \hat{c}_j \mathbf{u}_j\|_\infty^{-1}$ , where  $\hat{\mathbf{c}} := [\hat{c}_j : j \in \mathbb{N}_m]$  is a solution of the minimization problem (20) with  $\mathbf{y}$  and  $\mathcal{B}_* = c_0$ .*

## 5. Fixed-Point Approach for MNI

The solution method established in section 4 for the MNI problem in a Banach space with non-smooth pre-dual space provides a foundation for further development of implementable schemes to find its solution by determining the coefficients  $c_j$  which appear in the solution representations. Specifically, using the solution representation described in Theorem 25 to find a solution of problem (3) requires to solve the minimization problem (20) and to verify the inclusion relation. Both of these steps involve solving inclusion relations. It is not computationally convenient to solve an inclusion relation, especially, when the set involved in the inclusion is described by sophisticated equations and/or inequalities. It requires further investigation to develop computationally efficient schemes based on the theoretical results that we have obtained.

In this section, we take a different point of view to develop a fixed-point approach for the MNI problem in a Banach space. Specifically, we reformulate problem (3) as an unconstrained minimization problem, and then re-express its solution as a fixed-point of a nonlinear map defined via the proximity operator of functions involved in the problem. The resulting fixed-point equations can be solved efficiently by iteration schemes. The reformulation will be done by using the fact that an inclusion involving subdifferential of a convex function can be converted to a fixed-point equation defined by the proximity operator of the function. The fixed-point formulation provides a sound basis for algorithmic development for numerical solutions of the problem. In particular, when  $\mathcal{B}$  is the special Banach space  $\ell_1(\mathbb{N})$ , we develop an implementable fixed-point equation for finding a solution of this problem. Proofs of the results in this section will be included in Appendix E.

We now formulate fixed-point equations for a solution of problem (3) in a general Banach space. We first reformulate problem (3) as an equivalent unconstrained minimization problem. Suppose that  $\mathcal{B}$  is a real Banach space with the dual space  $\mathcal{B}^*$  and  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , are linearly independent. Let  $\mathcal{L}$  be defined by (1) and  $\mathcal{L}^*$  its adjoint operator. For a given vector  $\mathbf{y} \in \mathbb{R}^m$ , we define the indicator function  $\iota_{\mathbf{y}} : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  of  $\mathbf{y}$  at

$\mathbf{c} \in \mathbb{R}^m$  as

$$\iota_{\mathbf{y}}(\mathbf{c}) := \begin{cases} 0, & \text{if } \mathbf{c} = \mathbf{y}, \\ +\infty, & \text{if } \mathbf{c} \neq \mathbf{y}. \end{cases} \quad (22)$$

Note that the indicator function  $\iota_{\mathbf{y}}$  is convex but not continuous at  $\mathbf{c} = \mathbf{y}$ . Using this function, problem (3), a constrained minimization problem, is rewritten as an equivalent unconstrained one. We state this result in the next lemma for convenient reference.

**Lemma 33** *If for a given  $\mathbf{y} \in \mathbb{R}^m$ , the indicator function  $\iota_{\mathbf{y}}$  is defined by (22), then problem (3) with  $\mathbf{y}$  is equivalent to the unconstrained minimization problem*

$$\inf\{\|f\|_{\mathcal{B}} + \iota_{\mathbf{y}}(\mathcal{L}(f)) : f \in \mathcal{B}\}. \quad (23)$$

**Proof** By the definition of the indicator function, the infimum in (23) will be assumed at an element  $f \in \mathcal{B}$  such that  $\mathcal{L}(f) = \mathbf{y}$ . Thus, the minimization problem (23) can be rewritten as  $\inf\{\|f\|_{\mathcal{B}} : f \in \mathcal{B}, \mathcal{L}(f) = \mathbf{y}\}$ , which coincides with problem (3) with  $\mathbf{y}$ .  $\blacksquare$

We characterize a solution of problem (23) in terms of fixed-point equations. To this end, we need the notion of the proximity operator on both spaces  $\mathbb{R}^m$  and  $\mathcal{B}$ . We begin with reviewing the proximity operator on  $\mathbb{R}^m$  which was originally introduced in (Moreau, 1962). Let  $\psi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  be a convex function such that  $\text{dom}(\psi) := \{\mathbf{c} \in \mathbb{R}^m : \psi(\mathbf{c}) < +\infty\} \neq \emptyset$ . The proximity operator  $\text{prox}_{\psi} : \mathbb{R}^m \rightarrow \mathbb{R}^m$  of  $\psi$  is defined for  $\mathbf{a} \in \mathbb{R}^m$  by

$$\text{prox}_{\psi}(\mathbf{a}) := \arg \inf \left\{ \frac{1}{2} \|\mathbf{a} - \mathbf{c}\|_{\mathbb{R}^m}^2 + \psi(\mathbf{c}) : \mathbf{c} \in \mathbb{R}^m \right\}. \quad (24)$$

The proximity operator of a convex function in an infinite dimensional Hilbert space may be found in Bauschke and Combettes (2011). We now define the proximity operator of a convex function in a Banach space  $\mathcal{B}$ . This requires the availability of a Hilbert space and a linear map between it and the Banach space  $\mathcal{B}$ . Suppose that  $\mathcal{H}$  is a Hilbert space,  $\mathcal{T} : \mathcal{B} \rightarrow \mathcal{H}$  is a bounded linear operator and  $\mathcal{T}^*$  is its adjoint operator from  $\mathcal{H}$  to  $\mathcal{B}^*$ . The proximity operator  $\text{prox}_{\psi, \mathcal{H}, \mathcal{T}} : \mathcal{B} \rightarrow \mathcal{B}$  of a convex function  $\psi : \mathcal{B} \rightarrow \mathbb{R} \cup \{+\infty\}$  with respect to  $\mathcal{H}$  and  $\mathcal{T}$  is defined by

$$\text{prox}_{\psi, \mathcal{H}, \mathcal{T}}(f) := \arg \inf \left\{ \frac{1}{2} \|\mathcal{T}(f - h)\|_{\mathcal{H}}^2 + \psi(h) : h \in \mathcal{B} \right\}, \quad \text{for all } f \in \mathcal{B}. \quad (25)$$

The proximity operator  $\text{prox}_{\psi}$  defined by (24) is a special case of the definition (25). Specifically, let  $\mathcal{B}$  be the Euclidean space  $\mathbb{R}^m$  with a norm  $\|\cdot\|$ . If we choose  $\mathcal{H} := \mathbb{R}^m$  with the Euclidean norm  $\|\cdot\|_{\mathbb{R}^m}$  and  $\mathcal{T}$  as the identity operator from  $\mathbb{R}^m$  with the norm  $\|\cdot\|$  to  $\mathbb{R}^m$  with the Euclidean norm  $\|\cdot\|_{\mathbb{R}^m}$ , the proximity operator  $\text{prox}_{\psi, \mathcal{H}, \mathcal{T}}$  reduces to  $\text{prox}_{\psi}$ .

We also need the notion of the conjugate function of a convex function to develop a characterization for the solution of problem (23) in terms of fixed-point equations. The conjugate function of a convex function  $\psi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  is defined as  $\psi^*(\mathbf{c}) := \sup\{\langle \mathbf{a}, \mathbf{c} \rangle_{\mathbb{R}^m} - \psi(\mathbf{a}) : \mathbf{a} \in \mathbb{R}^m\}$ , for all  $\mathbf{c} \in \mathbb{R}^m$ . There is a relation between the subdifferential of a convex function and that of its conjugate function. Specifically, if  $\psi$  is a convex function on  $\mathbb{R}^m$ , then for all  $\mathbf{a} \in \text{dom}(\psi)$  and all  $\mathbf{b} \in \text{dom}(\psi^*)$  there holds

$$\mathbf{a} \in \partial\psi^*(\mathbf{b}) \quad \text{if and only if} \quad \mathbf{b} \in \partial\psi(\mathbf{a}). \quad (26)$$

This leads to the relation between the proximity operators of  $\psi$  and  $\psi^*$ , that is,  $\text{prox}_\psi = \mathbf{I} - \text{prox}_{\psi^*}$ , where  $\mathbf{I}$  denotes the  $m \times m$  identity matrix. As an example, the conjugate function  $\iota_{\mathbf{y}}^*$  of the indicator function  $\iota_{\mathbf{y}}$  has the form  $\iota_{\mathbf{y}}^*(\mathbf{c}) := \langle \mathbf{y}, \mathbf{c} \rangle_{\mathbb{R}^m}$ , for all  $\mathbf{c} \in \mathbb{R}^m$ .

We turn to characterizing a solution of problem (23) as a fixed-point of a nonlinear map defined via the proximity operators of the norm  $\|\cdot\|_{\mathcal{B}}$  and the conjugate function  $\iota_{\mathbf{y}}^*$ . It is proved by the chain rule for the subdifferential and the relations between the proximity operator of a convex function and its subdifferential, which are all stated in Appendix E. For convenience, we set  $\mathcal{V} := \text{span} \{\nu_j : j \in \mathbb{N}_m\}$ .

**Theorem 34** *Suppose that Assumption (A1) holds and  $\mathbf{y} \in \mathbb{R}^m$ . Let  $\mathcal{L}$  and  $\iota_{\mathbf{y}}$  be defined by (1) and (22), respectively,  $\mathcal{V}$  be defined as above and  $\mathcal{L}^*$  be the adjoint operator of  $\mathcal{L}$ . Let  $\mathcal{H}$  be a Hilbert space and  $\mathcal{T}$  a bounded linear operator from  $\mathcal{B}$  to  $\mathcal{H}$  with the adjoint operator  $\mathcal{T}^*$ . If  $\mathcal{T}^*\mathcal{T}$  is a one-to-one mapping from the inverse image of  $\mathcal{V}$  onto  $\mathcal{V}$ , then  $\hat{f} \in \mathcal{B}$  is a solution of problem (23) with  $\mathbf{y}$  if and only if there exists  $\mathbf{c} \in \mathbb{R}^m$  such that*

$$\mathbf{c} = \text{prox}_{\iota_{\mathbf{y}}^*}(\mathbf{c} + \mathcal{L}(\hat{f})) \quad \text{and} \quad \hat{f} = \text{prox}_{\|\cdot\|_{\mathcal{B}, \mathcal{H}, \mathcal{T}}}(\hat{f} - (\mathcal{T}^*\mathcal{T})^{-1}\mathcal{L}^*(\mathbf{c})). \quad (27)$$

Theorem 25 shows that when the Banach space  $\mathcal{B}$  has a non-smooth pre-dual space  $\mathcal{B}_*$  and  $\nu_j \in \mathcal{B}_*$ ,  $j \in \mathbb{N}_m$ , the coefficients  $c_j$ ,  $j \in \mathbb{N}_m$ , appearing in Theorem 12 can be obtained by solving the finite dimensional minimization problem (20). A solution of (20) can be alternatively characterized via fixed-point equations. We next present this result.

**Theorem 35** *Suppose that Assumption (A2) holds and  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$ . Let  $\mathcal{L}$  and  $\iota_{\mathbf{y}}$  be defined by (1) and (22), respectively,  $\mathcal{V}$  be defined earlier and  $\mathcal{L}^*$  be the adjoint operator of  $\mathcal{L}$ . Let  $\mathcal{H}$  be a Hilbert space and  $\mathcal{T}$  a bounded linear operator from  $\mathcal{B}$  to  $\mathcal{H}$  with the adjoint operator  $\mathcal{T}^*$ . If  $\mathcal{T}^*\mathcal{T}$  is a one-to-one mapping from the inverse image of  $\mathcal{V}$  onto  $\mathcal{V}$ , then  $\hat{\mathbf{c}} \in \mathbb{R}^m$  is a solution of problem (20) with  $\mathbf{y}$  if and only if there exists  $\hat{f} \in \mathcal{B}$  such that the pair  $\hat{f}$  and  $\hat{\mathbf{c}} := -\|\hat{f}\|_{\mathcal{B}}\hat{\mathbf{c}}$  satisfies the fixed-point equations (27).*

Theorem 34 shows that solving problem (3) can be done by solving fixed-point equations (27). These two fixed-point equations are coupled together and they have to be solved simultaneously by iteration. In general, the second fixed-point equation in (27) is of infinite dimension, which requires further investigation to reduce it to a finite dimensional fixed-point equation. We demonstrate this point by considering the case when  $\mathcal{B} = \ell_1(\mathbb{N})$  whose special property will enable us to reduce the corresponding fixed-point equation to a finite dimensional one.

Next, we establish a fixed-point characterization for a solution of problem (3) in the special Banach space  $\ell_1(\mathbb{N})$ . We are especially interested in showing how the fixed-point equations (27) of infinite dimension is reduced to equivalent finite dimensional fixed-point equations.

We first derive the proximity operator of convex functions on  $\ell_1(\mathbb{N})$ . According to definition (25), we need to choose an appropriate Hilbert space  $\mathcal{H}$  and the operator  $\mathcal{T} : \ell_1(\mathbb{N}) \rightarrow \mathcal{H}$ . Noting that there hold the inclusion relations  $\ell_1(\mathbb{N}) \subset \ell_2(\mathbb{N}) \subset c_0$ , we define the embedding operator  $\mathcal{T}_0 : \ell_1(\mathbb{N}) \rightarrow \ell_2(\mathbb{N})$  as

$$\mathcal{T}_0(\mathbf{x}) := \mathbf{x}, \quad \text{for all } \mathbf{x} \in \ell_1(\mathbb{N}). \quad (28)$$

Since there holds  $\|\mathbf{x}\|_2 \leq \|\mathbf{x}\|_1$ , for all  $\mathbf{x} \in \ell_1(\mathbb{N})$ , we conclude that  $\mathcal{T}_0$  is bounded. We next express the adjoint operator  $\mathcal{T}_0^*$  of  $\mathcal{T}_0$ . There holds for all  $\mathbf{u} \in \ell_2(\mathbb{N})$  and all  $\mathbf{x} \in \ell_1(\mathbb{N})$  that  $\langle \mathcal{T}_0^*(\mathbf{u}), \mathbf{x} \rangle_{\ell_1} = \langle \mathbf{u}, \mathcal{T}_0(\mathbf{x}) \rangle_{\ell_2} = \sum_{j \in \mathbb{N}} u_j x_j$ . This yields that  $\mathcal{T}_0^* : \ell_2(\mathbb{N}) \rightarrow \ell_\infty(\mathbb{N})$  has the form  $\mathcal{T}_0^*(\mathbf{u}) = \mathbf{u}$ , for all  $\mathbf{u} \in \ell_2(\mathbb{N})$ . By choosing  $\mathcal{H} := \ell_2(\mathbb{N})$  and  $\mathcal{T} := \mathcal{T}_0$  defined by (28), the proximity operator  $\text{prox}_{\psi, \ell_2(\mathbb{N}), \mathcal{T}_0} : \ell_1(\mathbb{N}) \rightarrow \ell_1(\mathbb{N})$  of a convex function  $\psi : \ell_1(\mathbb{N}) \rightarrow \mathbb{R} \cup \{+\infty\}$  defined by (25) has the form

$$\text{prox}_{\psi, \ell_2(\mathbb{N}), \mathcal{T}_0}(\mathbf{x}) = \arg \inf \left\{ \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 + \psi(\mathbf{z}) : \mathbf{z} \in \ell_1(\mathbb{N}) \right\}, \quad \text{for all } \mathbf{x} \in \ell_1(\mathbb{N}). \quad (29)$$

We can explicitly calculate the proximity operator (29) of the norm  $\psi := \|\cdot\|_1$  of  $\ell_1(\mathbb{N})$  and its subdifferential. Specifically, for each  $\mathbf{x} := (x_j : j \in \mathbb{N}) \in \ell_1(\mathbb{N})$  there holds

$$\partial \|\cdot\|_1(\mathbf{x}) = \{\mathbf{u} \in \ell_\infty(\mathbb{N}) : \mathbf{u} = (u_j : j \in \mathbb{N}), u_j \in \partial |\cdot|(x_j), j \in \mathbb{N}\},$$

where for each  $x \in \mathbb{R}$ ,  $\partial |\cdot|(x) := \text{sign}(x)$ , if  $x \neq 0$ , and  $\partial |\cdot|(x) := [-1, 1]$ , if  $x = 0$ , and

$$\text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0}(\mathbf{x}) = (\max\{|x_j| - 1, 0\} \text{sign}(x_j) : j \in \mathbb{N}). \quad (30)$$

We now turn to solving the minimization problem (23) in the case that  $\mathcal{B} := \ell_1(\mathbb{N})$ :

$$\inf \{ \|\mathbf{x}\|_1 + \iota_{\mathcal{Y}}(\mathcal{L}(\mathbf{x})) : \mathbf{x} \in \ell_1(\mathbb{N}) \}, \quad (31)$$

where  $\mathcal{L}$  is defined by (14) with  $\mathbf{u}_j \in c_0$ ,  $j \in \mathbb{N}_m$ . The solution of the minimization problem (31) can be characterized as a fixed-point of a map defined on a finite dimensional space. This benefits from an important property of the space  $c_0$ . Specifically, for each  $\mathbf{u} \in c_0$ , since  $u_j$  tends to zero as  $j \rightarrow +\infty$ ,  $\mathbf{u}$  attains its norm  $\|\mathbf{u}\|_\infty$  on the finite index set  $\mathbb{N}(\mathbf{u})$ . By virtue of this property, we introduce a truncation operator. We denote by  $c_c$  the space of all real sequences on  $\mathbb{N}$  having at most a finite number of nonzero components. Clearly, we have that  $c_c \subset c_0$ . For each  $\mathbf{x} \in c_c$ , the support of  $\mathbf{x}$ , denoted by  $\text{supp}(\mathbf{x})$ , is defined to be the index set on which  $\mathbf{x}$  is nonzero. We define the truncation operator  $\mathcal{S} : c_0 \rightarrow c_c$  as  $\mathcal{S}(\mathbf{u}) := (\tilde{u}_j : j \in \mathbb{N})$  with  $\tilde{u}_j := u_j$ , if  $j \in \mathbb{N}(\mathbf{u})$ , and 0 otherwise. Clearly, we have that  $\|\mathbf{u}\|_\infty = \|\mathcal{S}(\mathbf{u})\|_\infty$ , for all  $\mathbf{u} \in c_0$ .

We are now ready to characterize the solution of the minimization problem (31) by fixed-point equations. We let  $\mathcal{L}$ ,  $\iota_{\mathcal{Y}}$  and  $\mathcal{T}_0$  be defined by (14), (22) and (28), respectively, and  $\mathcal{S}$  be the truncation operator.

**Theorem 36** *Suppose that  $\mathbf{u}_j \in c_0$ ,  $j \in \mathbb{N}_m$ , are linearly independent and  $\mathbf{y} \in \mathbb{R}^m$ . Then  $\hat{\mathbf{x}} \in \ell_1(\mathbb{N})$  is a solution of the minimization problem (31) with  $\mathbf{y}$  if and only if there exists  $\mathbf{c} \in \mathbb{R}^m$  such that*

$$\mathbf{c} = \text{prox}_{\iota_{\mathcal{Y}}}(\mathbf{c} + \mathcal{L}(\hat{\mathbf{x}})) \quad \text{and} \quad \hat{\mathbf{x}} = \text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0}(\hat{\mathbf{x}} - \mathcal{S}\mathcal{L}^*(\mathbf{c})). \quad (32)$$

Following Theorem 36, we can give a characterization by fixed-point equations for the solution of the dual problem  $\inf \{ \|\mathcal{L}^*(\mathbf{c})\|_\infty : \langle \mathbf{c}, \mathbf{y} \rangle_{\mathbb{R}^m} = 1, \mathbf{c} \in \mathbb{R}^m \}$  of the MNI problem in the space  $\ell_1(\mathbb{N})$ .

**Theorem 37** *Suppose that  $\mathbf{u}_j \in c_0$ ,  $j \in \mathbb{N}_m$ , are linearly independent and  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$ . Then  $\hat{\mathbf{c}} \in \mathbb{R}^m$  is a solution of the dual problem with  $\mathbf{y}$  if and only if there exists  $\hat{\mathbf{x}} \in \ell_1(\mathbb{N})$  such that the pair  $\hat{\mathbf{x}}$  and  $\mathbf{c} := -\|\hat{\mathbf{x}}\|_1 \hat{\mathbf{c}}$  satisfy the fixed-point equations (32).*

The fixed-points equations (32) appearing in both Theorems 36 and 37 are in fact of finite dimension. We unfold this fact in the remaining part of this section. It is convenient to write the fixed-point equations (32) in a compact form. To this end, we stack the vector  $\mathbf{c} \in \mathbb{R}^m$  on the top of finite dimensional vector  $\hat{\mathbf{x}}$  to form a new vector  $\mathbf{s} := \begin{bmatrix} \mathbf{c} \\ \hat{\mathbf{x}} \end{bmatrix}$ . We also introduce two matrices of operators by

$$\mathcal{P} := \begin{bmatrix} \text{prox}_{\iota_{\mathbf{y}}^*} \\ \text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0} \end{bmatrix} \quad \text{and} \quad \mathcal{R} := \begin{bmatrix} \mathcal{I} & \mathcal{L} \\ -\mathcal{S}\mathcal{L}^* & \mathcal{I} \end{bmatrix}. \quad (33)$$

In the above notion, we rewrite equations (32) in the following compact form  $\mathbf{s} = (\mathcal{P} \circ \mathcal{R})(\mathbf{s})$ . The following theorem implies that this fixed-point equation (or equivalently the system of the fixed-points equations (32)) is of finite dimension.

**Theorem 38** *If operators  $\mathcal{P}$  and  $\mathcal{R}$  are defined as in (33), then  $\mathcal{P} \circ \mathcal{R}$  is an operator from  $(\mathbb{R}^m, \ell_1(\mathbb{N}))$  to  $(\mathbb{R}^m, c_c)$  and its fixed-point  $\mathbf{s} = \begin{bmatrix} \mathbf{c} \\ \hat{\mathbf{x}} \end{bmatrix} \in (\mathbb{R}^m, \ell_1(\mathbb{N}))$  satisfies  $\hat{\mathbf{x}} \in c_c$  and  $\text{supp}(\hat{\mathbf{x}}) \subseteq \text{supp}(\mathcal{S}(\mathcal{L}^*(\mathbf{c})))$ .*

A solution of problem (3) with  $\mathcal{B} := \ell_1(\mathbb{N})$  guaranteed by Theorems 36 and 38 has an additional property.

**Remark 39** *Each solution  $\hat{\mathbf{x}} \in \ell_1(\mathbb{N})$  of problem (3) with  $\mathcal{B} := \ell_1(\mathbb{N})$  together with  $\mathbf{c} \in \mathbb{R}^m$  satisfying the fixed-point equations (32) is of finite dimension, that is, it satisfies  $\hat{\mathbf{x}} \in c_c$  and  $\text{supp}(\hat{\mathbf{x}}) \subseteq \text{supp}(\mathcal{S}(\mathcal{L}^*(\mathbf{c})))$ .*

Theorem 36 reveals that to solve problem (3) with  $\mathcal{B} := \ell_1(\mathbb{N})$ , it suffices to find a solution of the fixed-point equations (32) by iterative algorithms designed based on these fixed-point equations. A remarkable fact is that according to Theorem 38, the fixed-point equations in (32) are both of *finite* dimension. Therefore, solving the infinitely dimensional problem (3) with  $\mathcal{B} := \ell_1(\mathbb{N})$  reduces to finding a fixed-point of a nonlinear map defined on a finite dimensional space.

To develop efficient iterative algorithms with convergence guaranteed based on these fixed-point equations, we need to consider additional issues: The first issue is the computation of the proximity operators of the two functions involved in the fixed-point equations. Moreover, the direct iteration from (32) may not lead to convergent algorithms. One needs to reformulate these fixed-point equations to equivalent ones guided by the theory of firmly non-expansive maps. This is the second issue. The third issue is how convergence of the resulting convergent iterative schemes can be accelerated by introducing some parameters or matrices.

We now address the first issue. Note that the closed-form formula for the proximity operator  $\text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0}$  has been given in (30). We now present the proximity operator of  $\iota_{\mathbf{y}}^*$  below. Clearly, by the definition of the indicator function  $\iota_{\mathbf{y}}$ , its proximity operator has the form  $\text{prox}_{\iota_{\mathbf{y}}}(\mathbf{a}) := \mathbf{y}$  for all  $\mathbf{a} \in \mathbb{R}^m$ . Then the relation between the proximity operators of  $\iota_{\mathbf{y}}$  and  $\iota_{\mathbf{y}}^*$  leads to closed-form formula of the proximity operator  $\text{prox}_{\iota_{\mathbf{y}}^*}$  as  $\text{prox}_{\iota_{\mathbf{y}}^*}(\mathbf{a}) = \mathbf{a} - \mathbf{y}$ , for all  $\mathbf{a} \in \mathbb{R}^m$ . The two closed-form formulas enable us to implement the iteration efficiently.



We next discuss the second issue. Since the equations (32) are represented in the equivalent compact form, one may define the Picard iteration based on the compact form fixed-point equation to find the fixed-point  $\mathbf{s}$ , that is  $\mathbf{s}^{k+1} = (\mathcal{P} \circ \mathcal{R})(\mathbf{s}^k)$ , for  $k = 0, 1, \dots$ . When it converges, the Picard sequence  $\mathbf{s}^k$ ,  $k = 0, 1, \dots$ , generated by the Picard iteration, converges to a fixed-point of the map  $\mathcal{P} \circ \mathcal{R}$ , which gives a solution of problem (3). However, convergence of the Picard sequence is not guaranteed. Normally, we need to reformulate the fixed-point equation by appropriately split the matrix  $\mathcal{R}$  guided by the theory of the non-expansive map. That is, we will construct from the map  $\mathcal{P} \circ \mathcal{R}$  a non-expansive map  $\mathcal{M}$  which has the same fixed-point set as  $\mathcal{P} \circ \mathcal{R}$ , so that the Picard sequence of the new map  $\mathcal{M}$  converges to a fixed-point of  $\mathcal{M}$ , guaranteed by its non-expansiveness. Interested readers are referred to (Li et al., 2015) for further algorithmic development along this line. We will address this issue together with other computational issues in a different occasion.

## 6. Regularization Problem and its Connection with MNI

We now consider regularization problems. In the remaining part of this paper, the term “regularization problem” will refer to both regularized learning and other semi-discrete inverse problems unless stated otherwise. Regularization problems are closely related to MNI problems. We shall translate the representer theorems obtained in section 3 for solutions of MNI problems to those of regularization problems. Since the regularization problem in a general Banach space is described as an infinity dimensional minimization problem, we first comment on the existence of a solution of the problem following general results regarding the existence of a solution of an infinity dimensional minimization problem. Moreover, we establish an intrinsic connection between the regularization problem and the MNI problem. Specifically, we shall show that there always exists a solution of the regularization problem which is also a solution of the MNI problem with specific data.

We first describe the regularization problem in a Banach space, review its background and several examples of practical importance, and establish existence of its solution under a rather mild condition. We begin with describing the regularization problem under investigation. Let  $\mathcal{B}$  be a real Banach space with the dual space  $\mathcal{B}^*$ . Suppose that a set of linearly independent functionals  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , is given and operator  $\mathcal{L} : \mathcal{B} \rightarrow \mathbb{R}^m$  is defined by equation (1). Learning a target element in  $\mathcal{B}$  from the given set of sampled data  $\{(\nu_j, y_j) : j \in \mathbb{N}_m\}$  consists of solving the following first kind operator equation

$$\mathcal{L}(f) = \mathbf{y} \tag{34}$$

for  $f \in \mathcal{B}$ , where  $\mathbf{y} := [y_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$ . Equation (34) is a typical ill-posed problem. That is, the inverse of  $\mathcal{L}$  is not bounded. A commonly used approach to address the ill-posedness of (34) is regularization. Specifically, we define a data fidelity term  $\mathcal{Q}_{\mathbf{y}}(\mathcal{L}(f))$  from (34) by using a loss function  $\mathcal{Q}_{\mathbf{y}} : \mathbb{R}^m \rightarrow \mathbb{R}_+$ , and solve the minimization problem

$$\inf\{\mathcal{Q}_{\mathbf{y}}(\mathcal{L}(f)) + \lambda\varphi(\|f\|_{\mathcal{B}}) : f \in \mathcal{B}\}, \tag{35}$$

where  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a regularizer and  $\lambda$  is a positive regularization parameter.

The regularization problem (35) appears in many applied areas. We present several examples of the loss function and the regularizer that are used frequently in applications.

In machine learning, classical regularization network and support vector machines for both classification and regression are reformulated as (35) (Evgeniou et al., 2000; Schölkopf and Smola, 2002). Specifically, if for  $\mathbf{y} \in \mathbb{R}^m$  the loss function is chosen as

$$Q_{\mathbf{y}}(\mathbf{z}) := \|\mathbf{z} - \mathbf{y}\|_{\mathbb{R}^m}^2, \quad \mathbf{z} \in \mathbb{R}^m, \quad (36)$$

and the quadratic regularizer as  $\varphi(t) := t^2$ ,  $t \in \mathbb{R}_+$ , then the regularization problem (35) reduces to the regularization networks. The support vector machine regression has the form (35) with the loss function being chosen as

$$Q_{\mathbf{y}}(\mathbf{z}) := \sum_{j \in \mathbb{N}_m} \max\{|y_j - z_j| - \epsilon, 0\}, \quad \mathbf{z} := [z_j : j \in \mathbb{N}_m] \in \mathbb{R}^m, \quad (37)$$

where  $\epsilon$  is a positive constant and the regularizer as the quadratic function. If the loss function is chosen as

$$Q_{\mathbf{y}}(\mathbf{z}) := \sum_{j \in \mathbb{N}_m} \max\{1 - y_j z_j, 0\}, \quad \mathbf{z} := [z_j : j \in \mathbb{N}_m] \in \{-1, 1\}^m, \quad (38)$$

for  $\mathbf{y} := [y_j : j \in \mathbb{N}_m] \in \{-1, 1\}^m$  and the regularizer as the quadratic function, the regularization problem (35) describes the support vector machine classification. Moreover,  $\ell_1$  support vector machine regression and classification (Li et al., 2018, 2019; Schölkopf and Smola, 2002; Zhu et al., 2004) are formulated as (35) with the loss function (37) and (38), respectively, and the linear regularizer  $\varphi(t) := t$ ,  $t \in \mathbb{R}_+$ . The Lasso regularized model (Tibshirani, 1996; Zhao and Yu, 2006) is also formulated as (35) with the loss function as (36) and the regularizer as the linear function with an appropriate choice of the Banach space. Another example concerns the  $l_p$ -norm regularization (Zhang, 2002) in which the regularizer is chosen as  $\varphi(t) := t^p$ ,  $t \in \mathbb{R}_+$ .

Most data science problems are described as SDIPs (Daubechies et al., 2004; Wendland, 2005). Such inverse problems cover many important application areas including image restoration (Cai et al., 2012; Lu et al., 2010) and medical imaging (Chen et al., 2020; Jiang et al., 2019). SDIPs often solved by regularization methods (Chen et al., 2015; Chen et al., 2008) are formulated in the form (35) with appropriate choices of the loss function and regularizer. The form of the loss function is normally determined by types of noise contaminated in given data.

We now consider the existence of a solution of the regularization problem (35). By using arguments similar to those used in the proof of the existence of solutions of problem (3), we can get the existence of solutions of (35) under the conditions that  $\mathcal{B}$  has a pre-dual space  $\mathcal{B}_*$  and  $\nu_j \in \mathcal{B}_*$ ,  $j \in \mathbb{N}_m$ . To this end, we review a few useful properties of functions appearing in (35). A function  $\mathcal{T}$  mapping from a topological space  $\mathcal{X}$  into  $\mathbb{R}$  is said to be lower semi-continuous if  $\mathcal{T}(f) \leq \liminf_{\alpha} \mathcal{T}(f_{\alpha})$ , whenever  $f_{\alpha}$ ,  $\alpha \in I$ , for some index set  $I$  is a net in  $\mathcal{X}$  converging to some element  $f \in \mathcal{X}$ . The notion of weakly\* lower semi-continuous is defined accordingly under the weak\* topology. We say a function  $\mathcal{T}$  mapping from a normed space  $\mathcal{X}$  into  $\mathbb{R}$  is coercive if  $\lim_{\|x\| \rightarrow +\infty} \mathcal{T}(x) = \infty$ .

We establish in the following proposition a sufficient condition which ensures the existence of a solution of the regularization problem (35). Its complete proof is included in Appendix F.

**Proposition 40** *Suppose that  $\mathbf{y} \in \mathbb{R}^m$ , both  $\mathcal{Q}_{\mathbf{y}} : \mathbb{R}^m \rightarrow \mathbb{R}_+$  and  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  are lower semi-continuous,  $\lambda > 0$  and moreover,  $\varphi$  is increasing and coercive. If  $\mathcal{B}$  is a Banach space having a pre-dual space  $\mathcal{B}_*$  and the functionals  $\nu_j, j \in \mathbb{N}_m$ , appearing in the definition of  $\mathcal{L}$  are in  $\mathcal{B}_*$ , then the regularization problem (35) has at least one solution.*

The reflexive Banach space is a special case of the Banach space having a pre-dual space. In this special case, we can also have the existence result for a solution of the regularization problem (35).

We now turn to investigating the connection between a solution of the regularization problem (35) and that of the MNI problem (3). Specifically, we shall show that if the regularizer is increasing then there always exists a solution of the regularization problem (35) which is also a solution of the MNI problem with specific data. Furthermore, if the regularizer is strictly increasing then every solution of the regularization problem (35) is also a solution of the MNI problem with specific data. Throughout the rest of this paper, we always assume that each of the two minimization problems has a solution without further mention. In particular, it is guaranteed by Proposition 40 that this assumption holds when  $\mathcal{B}$  has a pre-dual space  $\mathcal{B}_*$ ,  $\nu_j \in \mathcal{B}_*$ ,  $j \in \mathbb{N}_m$ ,  $\mathcal{Q}_{\mathbf{y}}$ ,  $\varphi$  are both lower semi-continuous and  $\varphi$  is increasing and coercive.

**Proposition 41** *Suppose that  $\mathcal{B}$  is a Banach space with the dual space  $\mathcal{B}^*$ ,  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , and  $\mathcal{L}$  is defined by (1). For a given  $\mathbf{y}_0 \in \mathbb{R}^m$ , let  $\mathcal{Q}_{\mathbf{y}_0} : \mathbb{R}^m \rightarrow \mathbb{R}_+$  be a loss function,  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be an increasing regularizer and  $\lambda > 0$ . Let  $\hat{f} \in \mathcal{B}$  be a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$ . Then the following statements hold true:*

- (i) *A solution  $\hat{g} \in \mathcal{B}$  of problem (3) with  $\mathbf{y} := \mathcal{L}(\hat{f})$  is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$ .*
- (ii) *If  $\varphi$  is strictly increasing, then  $\hat{f}$  is a solution of problem (3) with  $\mathbf{y} := \mathcal{L}(\hat{f})$ .*

Statement (ii) of Proposition 41 was claimed in Micchelli and Pontil (2004) without details of proof. We provide a complete proof for this statement in Appendix F for convenience of the readers.

## 7. Representer Theorems and Solution Methods for Regularization Problems

In this section, we establish representer theorems and solution methods for solutions of the regularization problem (35), with proofs included in Appendix G. Using the connection enacted in the last section between a solution of the MNI problem and that of the regularization problem, we first present both implicit and explicit representer theorems for a solution of the regularization problem (35). We then develop solution methods for solving the regularization problem. We present two types of solution methods: one based on the representer theorems and the other being direct methods. We also consider special cases and give special results for them. In particular, for the regularization problem in  $\ell_1(\mathbb{N})$ , we put forward a fixed-point formulation which serves as a basis for further development of efficient iterative algorithms for solving the problem. Although results to be presented in this section are parallel to those for MNI, they will provide a foundation for applications due

to wide utilizations of regularization problems in many areas. We will keep our presentation concise by skipping some details.

We first present representer theorems for the regularization problem (35). Recall that we have established several implicit representer theorems in Proposition 7 for solutions of the MNI problem (3). Through the connection between a solution of the regularization problem (35) and that of problem (3), in the next proposition we first translate the results in Proposition 7 originally for problem (3) to those for the regularization problem (35). In this section, we always assume that for a given  $\mathbf{y}_0 \in \mathbb{R}^m$ ,  $\mathcal{Q}_{\mathbf{y}_0} : \mathbb{R}^m \rightarrow \mathbb{R}_+$  is a loss function,  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  is a regularizer and  $\lambda > 0$ .

**Proposition 42** *Suppose that  $\mathcal{B}$  is a Banach space with the dual space  $\mathcal{B}^*$ ,  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ ,  $\mathbf{y}_0 \in \mathbb{R}^m$  and  $\mathcal{L}$  is defined by (1). If  $\varphi$  is increasing, then there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  satisfying the following conditions:*

- (i) *There exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that the linear functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  satisfying  $\|\nu\|_{\mathcal{B}^*} = 1$  and  $\langle \nu, f_0 \rangle_{\mathcal{B}} = \|f_0\|_{\mathcal{B}}$ .*
- (ii) *There exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that the linear functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  peaks at  $f_0$ , that is,  $\langle \nu, f_0 \rangle_{\mathcal{B}} = \|\nu\|_{\mathcal{B}^*} \|f_0\|_{\mathcal{B}}$ .*
- (iii) *There exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\sum_{j \in \mathbb{N}_m} c_j \nu_j \in \mathcal{J}(f_0)$ .*
- (iv) *There exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\sum_{j \in \mathbb{N}_m} c_j \nu_j \in \partial \|\cdot\|_{\mathcal{B}}(f_0)$ .*

*If  $\varphi$  is strictly increasing, then every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  satisfies the above conditions (i)-(iv).*

When the Banach space  $\mathcal{B}$  is smooth, we may get a representer theorem for a solution of the regularization problem (35) in a simple form. The desired result can be obtained by translating representer Theorem 8 for problem (3) to problem (35) through Proposition 41 and using the arguments similar to those in the proof of Proposition 42.

**Remark 43** *Suppose that  $\mathcal{B}$  is a smooth Banach space. If  $\varphi$  is increasing, then there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  such that  $\mathcal{G}(f_0) = \sum_{j \in \mathbb{N}_m} c_j \nu_j$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ . Moreover, if  $\varphi$  is strictly increasing, then every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  has the form  $\mathcal{G}(f_0) = \sum_{j \in \mathbb{N}_m} c_j \nu_j$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

We next develop explicit representer theorems for the regularization problem (35). These explicit representer theorems are obtained from the explicit representer theorems for problem (3) presented in section 3 in conjunction with Proposition 41. We first consider the case when a Banach space  $\mathcal{B}$  has the dual space  $\mathcal{B}^*$ . The following theorem results from Theorem 10.

**Theorem 44** *Suppose that  $\mathcal{B}$  is a Banach space with the dual space  $\mathcal{B}^*$ ,  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ ,  $\mathbf{y}_0 \in \mathbb{R}^m$  and  $\mathcal{L}$  is defined by (1).*

- (i) *If  $\varphi$  is increasing, then there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  such that  $f_0 \in \gamma \partial \|\cdot\|_{\mathcal{B}^*}(\sum_{j \in \mathbb{N}_m} c_j \nu_j)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , with  $\gamma := \|\sum_{j \in \mathbb{N}_m} c_j \nu_j\|_{\mathcal{B}^*}$ .*
- (ii) *If  $\varphi$  is strictly increasing, then every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  satisfies  $f_0 \in \gamma \partial \|\cdot\|_{\mathcal{B}^*}(\sum_{j \in \mathbb{N}_m} c_j \nu_j)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

When the Banach space  $\mathcal{B}$  has the smooth dual space  $\mathcal{B}^*$ , we have a special representer theorem for a solution of problem (35).

**Remark 45** *Suppose that the dual space  $\mathcal{B}^*$  of the Banach space  $\mathcal{B}$  is smooth. If  $\varphi$  is increasing, then there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  in the form  $f_0 = \gamma \mathcal{G}^*(\sum_{j \in \mathbb{N}_m} c_j \nu_j)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ . Moreover, if  $\varphi$  is strictly increasing, then every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  has the form  $f_0 = \gamma \mathcal{G}^*(\sum_{j \in \mathbb{N}_m} c_j \nu_j)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

Below, we derive representer theorems for a solution of problem (35) in the case when the Banach space  $\mathcal{B}$  has a pre-dual space  $\mathcal{B}_*$  and  $\nu_j \in \mathcal{B}_*$ ,  $j \in \mathbb{N}_m$ . We first obtain an explicit solution representation from Theorem 12.

**Theorem 46** *Suppose that  $\mathcal{B}$  is a Banach space having a pre-dual space  $\mathcal{B}_*$ ,  $\nu_j \in \mathcal{B}_*$ ,  $j \in \mathbb{N}_m$ ,  $\mathbf{y}_0 \in \mathbb{R}^m$  and  $\mathcal{L}$  is defined by (1).*

(i) *If  $\varphi$  is increasing, then there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  such that  $f_0 \in \rho \partial \| \cdot \|_{\mathcal{B}_*}(\sum_{j \in \mathbb{N}_m} c_j \nu_j)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , with  $\rho := \| \sum_{j \in \mathbb{N}_m} c_j \nu_j \|_{\mathcal{B}_*}$ .*

(ii) *If  $\varphi$  is strictly increasing, then every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  satisfies  $f_0 \in \rho \partial \| \cdot \|_{\mathcal{B}_*}(\sum_{j \in \mathbb{N}_m} c_j \nu_j)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

When the Banach space  $\mathcal{B}$  has a smooth pre-dual space  $\mathcal{B}_*$ , we have a special representer theorem for a solution of problem (35).

**Remark 47** *Suppose that the Banach space  $\mathcal{B}$  has a smooth pre-dual space  $\mathcal{B}_*$ . If  $\varphi$  is increasing, then there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  in the form  $f_0 = \rho \mathcal{G}_*(\sum_{j \in \mathbb{N}_m} c_j \nu_j)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ . Moreover, if  $\varphi$  is strictly increasing, then every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  has the form  $f_0 = \rho \mathcal{G}_*(\sum_{j \in \mathbb{N}_m} c_j \nu_j)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

Observing from the above representer theorems, the essence of the representer theorems is that the original optimization problem in an infinite dimensional space can be converted to one in a finite dimensional space. This benefits from the fact that the number of data points, used in the regularization problem, is finite.

We consider below representer theorems for several special cases of Banach spaces and present special results. Regularized learning was originally considered to learn a function in an RKHS from a finite number of point-evaluation functional data, that is,  $\nu_j := \delta_{x_j}$ ,  $j \in \mathbb{N}_m$ , where  $x_j, j \in \mathbb{N}_m$ , are finite points in an input set  $X$ . Suppose that  $\mathcal{H}$  is an RKHS on  $X$  with the reproducing kernel  $K$ . If for a given  $\mathbf{y}_0 \in \mathbb{R}^m$ ,  $\mathcal{Q}_{\mathbf{y}_0}$  and  $\varphi$  are continuous and convex and moreover,  $\varphi$  is strictly increasing and coercive, then there exists a unique solution  $f_0$  of problem (35) with  $\mathbf{y} := \mathbf{y}_0$ . Note that for each  $j \in \mathbb{N}_m$ ,  $K(x_j, \cdot)$  refers to a closed-form function representation of  $\nu_j := \delta_{x_j}$ . Since  $\varphi$  is strictly increasing, by Remark 45 with  $\mathcal{B} := \mathcal{H}$  and  $\nu_j := K(x_j, \cdot)$ ,  $j \in \mathbb{N}_m$ , we express  $f_0$  in the form  $f_0 = \| \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot) \|_{\mathcal{H}^*} \mathcal{G}^*(\sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot))$  for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ . Note that  $\mathcal{H}_* = \mathcal{H}$ . Substituting  $f = \|f\|_{\mathcal{H}} \mathcal{G}(f)$  with  $f := \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot)$  into the above form, we get the representation of  $f_0$  as  $f_0 = \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot)$ .

We next consider regularization problems in a functional reproducing kernel Hilbert space (FRKHS). Motivated by learning a function from a finite number of non-point-evaluation functional data, we introduce in Wang and Xu (2019) the notion of FRKHSs. Let  $\mathcal{H}$  be a Hilbert space and  $\mathcal{F}$  a family of linear functionals on  $\mathcal{H}$ . Space  $\mathcal{H}$  is called an FRKHS with respect to  $\mathcal{F}$  if the norm of  $\mathcal{H}$  is compatible with  $\mathcal{F}$  and each linear functional in  $\mathcal{F}$  is continuous on  $\mathcal{H}$ . An FRKHS is expected to admit a reproducing kernel, which reproduces the linear functionals defining the space. Specifically, for an FRKHS  $\mathcal{H}$  with respect to a family  $\mathcal{F} := \{\nu_\alpha : \alpha \in \Lambda\}$  of linear functionals, there exists a unique functional reproducing kernel  $K : \Lambda \rightarrow \mathcal{H}$  such that  $K(\alpha) \in \mathcal{H}$ , for all  $\alpha \in \Lambda$ , and  $\nu_\alpha(f) = \langle f, K(\alpha) \rangle_{\mathcal{H}}$ , for all  $f \in \mathcal{H}$  and for all  $\alpha \in \Lambda$ . The representer theorem for regularization problems in an FRKHS from a finite number of non-point-evaluation functional data  $\nu_{\alpha_j} := K(\alpha_j)$ ,  $j \in \mathbb{N}_m$ , can also be derived from Remark 45. It suffices to notice that for each  $\alpha \in \Lambda$ ,  $K(\alpha)$  is an explicit representation for the functionals  $\nu_\alpha$ . Specifically, if for a given  $\mathbf{y}_0 \in \mathbb{R}^m$ ,  $\mathcal{Q}_{\mathbf{y}_0}$  and  $\varphi$  are continuous and convex and moreover,  $\varphi$  is strictly increasing and coercive, then there exists a unique solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  and it has the form  $f_0 = \sum_{j \in \mathbb{N}_m} c_j K(\alpha_j)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .

We now turn to considering regularization problems in a uniformly Fréchet smooth and uniformly convex Banach space  $\mathcal{B}$ . Recall that in such a Banach space, the semi-inner-product may be taken as a substitute of the inner product in a Hilbert space. Accordingly, each continuous linear functional on  $\mathcal{B}$  can be represented by the dual element of a unique element in  $\mathcal{B}$ , which is defined via the semi-inner-product. In particular, the Gâteaux derivative of the norm  $\|\cdot\|_{\mathcal{B}}$  has the form (66).

Applying Remark 43 to the regularization problem (35) in a uniformly Fréchet smooth and uniformly convex Banach space, we get the representer theorem as follows. Note that in this case, the linearly independent functionals have the form  $\nu_j := g_j^\sharp$  for  $g_j \in \mathcal{B}$ ,  $j \in \mathbb{N}_m$ .

**Theorem 48** *Suppose that Assumption (A3) holds and  $\mathbf{y}_0 \in \mathbb{R}^m$ .*

- (i) *If  $\varphi$  is increasing, then there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  such that  $f_0^\sharp = \sum_{j \in \mathbb{N}_m} c_j g_j^\sharp$  for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*
- (ii) *If  $\varphi$  is strictly increasing, then every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  satisfies  $f_0^\sharp = \sum_{j \in \mathbb{N}_m} c_j g_j^\sharp$  for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

It is desirable to have a representation for  $f_0$  in addition to that for  $f_0^\sharp$ . Since the uniformly Fréchet smooth and uniformly convex Banach space  $\mathcal{B}$  has the smooth dual space  $\mathcal{B}^*$ , Remark 45 allows us to have a representation for  $f_0$  in such a Banach space  $\mathcal{B}$ .

**Theorem 49** *Suppose that Assumption (A3) holds and  $\mathbf{y}_0 \in \mathbb{R}^m$ .*

- (i) *If  $\varphi$  is increasing, then there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  such that  $f_0 = \left( \sum_{j \in \mathbb{N}_m} c_j g_j^\sharp \right)^\sharp$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*
- (ii) *If  $\varphi$  is strictly increasing, then every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  satisfies  $f_0 = \left( \sum_{j \in \mathbb{N}_m} c_j g_j^\sharp \right)^\sharp$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

We now apply Theorem 49 to problem (35) in  $\ell_p(\mathbb{N})$  with  $1 < p < +\infty$ , special uniformly Fréchet smooth and uniformly convex Banach spaces. Suppose that  $\mathbf{u}_j \in \ell_q(\mathbb{N})$ ,  $j \in \mathbb{N}_m$ ,

with  $1/p + 1/q = 1$ , and operator  $\mathcal{L}$  is defined by (13). If for a given  $\mathbf{y}_0 \in \mathbb{R}^m$ ,  $\mathcal{Q}_{\mathbf{y}_0}$  and  $\varphi$  are continuous and convex and moreover,  $\varphi$  is strictly increasing and coercive, then there exists a unique solution  $\hat{\mathbf{x}} := (\hat{x}_j : j \in \mathbb{N})$  of the regularization problem (35) in  $\ell_p(\mathbb{N})$  with  $\mathbf{y} = \mathbf{y}_0$ . If  $\hat{\mathbf{x}} \neq 0$ , then  $\hat{x}_j = u_j |u_j|^{q-2} / \|\mathbf{u}\|_q^{q-2}$ , where  $\mathbf{u} = (u_j : j \in \mathbb{N}) := \sum_{j \in \mathbb{N}_m} c_j \mathbf{u}_j$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .

The semi-inner-product RKBS is uniformly Fréchet smooth and uniformly convex. A regularization problem in such a space is considered to learn a function from the sample data produced by point-evaluation functionals  $\nu_j := \delta_{x_j}$ ,  $j \in \mathbb{N}_m$ , where  $x_j, j \in \mathbb{N}_m$ , are a finite number of points in an input set  $X$ . By the reproducing property, the dual element  $G(x_j, \cdot)^\sharp$  of  $G(x_j, \cdot)$  coincides exactly with the point-evaluation functional  $\delta_{x_j}$  for  $j \in \mathbb{N}_m$ . Applying Theorem 48 with  $g_j := G(x_j, \cdot)$ ,  $j \in \mathbb{N}_m$ , to the regularization problem (35) in a semi-inner-product RKBS, we get the representer theorem as follows.

**Corollary 50** *Suppose that  $\mathcal{B}$  is the semi-inner-product RKBS with the semi-inner-product reproducing kernel  $G$  and  $x_j \in X$ ,  $j \in \mathbb{N}_m$ . Let  $\mathbf{y}_0 \in \mathbb{R}^m$ .*

- (i) *If  $\varphi$  is increasing, then there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  such that  $f_0^\sharp = \sum_{j \in \mathbb{N}_m} c_j G(x_j, \cdot)^\sharp$  for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*
- (ii) *If  $\varphi$  is strictly increasing, then every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  satisfies  $f_0^\sharp = \sum_{j \in \mathbb{N}_m} c_j G(x_j, \cdot)^\sharp$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

An explicit representation for the solution of the regularization problem in the semi-inner-product RKBS  $\mathcal{B}$  can also be obtained. The desired result is an immediate consequence of Theorem 49 with  $g_j := G(x_j, \cdot)$ , for  $j \in \mathbb{N}_m$ .

**Theorem 51** *Suppose that  $\mathcal{B}$  is the semi-inner-product RKBS with the semi-inner-product reproducing kernel  $G$  and  $x_j \in X$ ,  $j \in \mathbb{N}_m$ . Let  $\mathbf{y}_0 \in \mathbb{R}^m$ .*

- (i) *If  $\varphi$  is increasing, then there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  such that  $f_0 = \left( \sum_{j \in \mathbb{N}_m} c_j G(x_j, \cdot) \right)^\sharp$  for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*
- (ii) *If  $\varphi$  is strictly increasing, then every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  satisfies  $f_0 = \left( \sum_{j \in \mathbb{N}_m} c_j G(x_j, \cdot) \right)^\sharp$  for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

We turn to the regularization problem in a right-sided RKBS  $\mathcal{B}$  with the right-sided reproducing kernel  $K$ . Note that the right-sided reproducing kernel  $K$  provides the closed-form function representation for the point-evaluation functionals. By Remark 43 with  $\nu_j := K(x_j, \cdot)$ ,  $j \in \mathbb{N}_m$ , we have the following representer theorem. In this case, the assumptions on  $\mathcal{Q}_{\mathbf{y}_0}$ ,  $\varphi$  and the right-sided RKBS ensure the existence and uniqueness of the solution of problem (35).

**Corollary 52** *Suppose that Assumption (A4) holds and  $\mathcal{B}$  is reflexive, strictly convex and smooth. If for a given  $\mathbf{y}_0 \in \mathbb{R}^m$ ,  $\mathcal{Q}_{\mathbf{y}_0}$  and  $\varphi$  are continuous and convex and moreover,  $\varphi$  is strictly increasing and coercive, then the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  has a unique solution  $f_0$  and it has the form  $\mathcal{G}(f_0) = \sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot)$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

It is known that the reflexive Banach space  $\mathcal{B}$  is strict convexity if and only if the dual space  $\mathcal{B}^*$  is smooth. Hence, by Remark 45 with  $\nu_j := K(x_j, \cdot)$ ,  $j \in \mathbb{N}_m$ , an explicit representation for  $f_0$  can be obtained as follows.

**Theorem 53** *Suppose that Assumption (A4) holds and  $\mathcal{B}$  is reflexive and strictly convex. If for a given  $\mathbf{y}_0 \in \mathbb{R}^m$ ,  $\mathcal{Q}_{\mathbf{y}_0}$  and  $\varphi$  are continuous and convex and moreover,  $\varphi$  is strictly increasing and coercive, then the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  has a unique solution  $f_0$  and it has the form  $f_0 = \sigma \mathcal{G}^*(\sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot))$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , where  $\sigma := \|\sum_{j \in \mathbb{N}_m} c_j K(x_j, \cdot)\|_{\mathcal{B}^*}$ .*

Finally, we specialize Theorem 46 to the regularization problem in the space  $\ell_1(\mathbb{N})$ .

**Theorem 54** *Suppose that  $\mathbf{u}_j \in c_0$ ,  $j \in \mathbb{N}_m$ ,  $\mathbf{y}_0 \in \mathbb{R}^m$  and operator  $\mathcal{L}$  is defined by (14).*

(i) *If  $\varphi$  is increasing, then there exists a solution  $\hat{\mathbf{x}} \in \ell_1(\mathbb{N})$  of the regularization problem (35) in  $\ell_1(\mathbb{N})$  with  $\mathbf{y} := \mathbf{y}_0$  such that  $\hat{\mathbf{x}} \in \|\mathbf{u}\|_\infty \text{co}(\mathcal{V}(\mathbf{u}))$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , where  $\mathbf{u} := \sum_{j \in \mathbb{N}_m} c_j \mathbf{u}_j$ .*

(ii) *If  $\varphi$  is strictly increasing, then every solution  $\hat{\mathbf{x}}$  of the regularization problem (35) in  $\ell_1(\mathbb{N})$  with  $\mathbf{y} := \mathbf{y}_0$  satisfies  $\hat{\mathbf{x}} \in \|\mathbf{u}\|_\infty \text{co}(\mathcal{V}(\mathbf{u}))$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ .*

The representer theorems presented above for the regularization problem give only forms of the solutions for the problem, the same as those for MNI developed in section 3. To provide solution methods for the regularization problem, one has to determine the coefficients  $c_j$  involved in the solution representations. Next, we develop approaches to determine these coefficients, leading to solution methods for solving the regularization problem. We also consider solving directly the regularization problem (35) by the fixed-point approach, which has been discussed in section 5 for the MNI problem.

We begin with considering the case that the Banach space  $\mathcal{B}$  has a smooth pre-dual space  $\mathcal{B}_*$  and  $\nu_j \in \mathcal{B}_*$ , for  $j \in \mathbb{N}_m$ . In this case, Remark 47 provides a simple and explicit representation for the solutions of the regularization problem (35). By employing this solution representation, problem (35) can be converted to a finite dimensional minimization problem about the coefficients appearing in the representation.

**Theorem 55** *Suppose that  $\mathcal{B}$  is a Banach space having a smooth pre-dual space  $\mathcal{B}_*$ , and  $\nu_j \in \mathcal{B}_*$ ,  $j \in \mathbb{N}_m$ . Let  $\mathcal{L}$  be the linear operator defined by (1) and  $\mathcal{L}^*$  be the adjoint operator. For a given  $\mathbf{y}_0 \in \mathbb{R}^m$ , let  $\mathcal{Q}_{\mathbf{y}_0} : \mathbb{R}^m \rightarrow \mathbb{R}_+$  be a loss function,  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a strictly increasing regularizer and  $\lambda > 0$ . Then*

$$f_0 := \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*} \mathcal{G}_*(\mathcal{L}^*(\hat{\mathbf{c}})), \quad \hat{\mathbf{c}} \in \mathbb{R}^m, \quad (39)$$

*is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  if and only if  $\hat{\mathbf{c}} \in \mathbb{R}^m$  is a solution of the minimization problem*

$$\inf\{\mathcal{Q}_{\mathbf{y}_0}(\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} \mathcal{L}(\mathcal{G}_*(\mathcal{L}^*(\mathbf{c})))) + \lambda \varphi(\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*}) : \mathbf{c} \in \mathbb{R}^m\}. \quad (40)$$

As a special case, we consider the regularization problem (35) in a Hilbert space  $\mathcal{H}$ , that is  $\mathcal{B} := \mathcal{H}$ . In this case, the regularizer  $\varphi$  has the form  $\varphi(t) := t^2$ ,  $t \in \mathbb{R}_+$ , which is strictly increasing on  $\mathbb{R}_+$ ,  $\mathcal{B}_* = \mathcal{H}$  and the linearly independent functionals  $\nu_j$ ,  $j \in \mathbb{N}_m$ , are identified with  $g_j \in \mathcal{H}$ .

**Corollary 56** *Suppose that  $\mathcal{H}$  is a Hilbert space and  $g_j \in \mathcal{H}$ ,  $j \in \mathbb{N}_m$ . Let  $\mathcal{L}$  be the linear operator defined by (1) with  $\nu_j := g_j$ ,  $j \in \mathbb{N}_m$ ,  $\mathcal{L}^*$  be the adjoint operator and  $\mathbf{G}$  be the*



*Gram matrix.* For a given  $\mathbf{y}_0 \in \mathbb{R}^m$ , let  $\mathcal{Q}_{\mathbf{y}_0} : \mathbb{R}^m \rightarrow \mathbb{R}_+$  be a loss function and  $\lambda > 0$ . Then  $f_0 := \sum_{j \in \mathbb{N}_m} \hat{c}_j g_j$  is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  if and only if  $\hat{\mathbf{c}} := [\hat{c}_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$  is a solution of the minimization problem

$$\inf\{\mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}\mathbf{c}) + \lambda \mathbf{c}^\top \mathbf{G}\mathbf{c} : \mathbf{c} \in \mathbb{R}^m\}. \quad (41)$$

Below, we discuss how to solve the finite dimensional minimization problem (41). Approaches that may be adopted to solve the problem (41) depend on the smoothness of the function  $\mathcal{Q}_{\mathbf{y}_0}$  appearing in the fidelity term of (41). When  $\mathcal{Q}_{\mathbf{y}_0}$  is differentiable, the minimization problem (41) may be solved by using standard methods such as the gradient descent method. When  $\mathcal{Q}_{\mathbf{y}_0}$  is not differentiable, standard methods for solving minimization problems are not applicable to problem (41) and it requires special treatment. We will pay a special attention to the case when  $\mathcal{Q}_{\mathbf{y}_0}$  is not differentiable.

We now consider solving the finite dimensional minimization problem (41). The following solution methods are described for both differentiable and non-differentiable loss functions.

**Remark 57** For a given  $\mathbf{y}_0 \in \mathbb{R}^m$ , let  $\mathcal{Q}_{\mathbf{y}_0}$  be a convex loss function. If  $\mathcal{Q}_{\mathbf{y}_0}$  is non-differentiable, then  $\hat{\mathbf{c}} \in \mathbb{R}^m$  is the unique solution of the minimization problem (41) if and only if  $\hat{\mathbf{c}}$  satisfies  $\hat{\mathbf{c}} = \frac{1}{-2\lambda} \text{prox}_{\mathcal{Q}_{\mathbf{y}_0}^*}(-2\lambda\hat{\mathbf{c}} + \mathbf{G}\hat{\mathbf{c}})$ .

In the case that the loss function is differentiable, the solution of the finite dimensional minimization problem (41) satisfies a system which usually is nonlinear. For the loss function with a special form, the nonlinear system reduces to a linear one.

**Remark 58** If the convex function  $\mathcal{Q}_{\mathbf{y}_0}$  is differentiable, then  $\hat{\mathbf{c}}$  is the unique solution of the minimization problem (41) if and only if  $\hat{\mathbf{c}}$  satisfies the system  $-2\lambda\hat{\mathbf{c}} = \nabla \mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}\hat{\mathbf{c}})$ . Particularly, if  $\mathcal{Q}_{\mathbf{y}_0}$  has the form (36), then the system reduces to the linear system  $(\mathbf{G} + \lambda\mathbf{I})\hat{\mathbf{c}} = \mathbf{y}_0$ .

Our second example concerns the regularization problem in a uniformly Fréchet smooth and uniformly convex Banach space  $\mathcal{B}$ . In such a space, there exists a unique semi-inner-product  $[\cdot, \cdot]_{\mathcal{B}}$  that induces the norm  $\|\cdot\|_{\mathcal{B}}$ . Moreover, for each  $\nu \in \mathcal{B}^*$ , there exists a unique  $g \in \mathcal{B}$  such that  $\nu = g^\sharp$ . Thus, in this case, the linear functional  $\nu_j$  appearing in (35) is identified with  $g_j^\sharp$ , for  $g_j \in \mathcal{B}$ ,  $j \in \mathbb{N}_m$ . With respect to the sequence  $g_j \in \mathcal{B}$ ,  $j \in \mathbb{N}_m$ , we introduce a nonlinear operator  $\mathbf{G}_{\text{s.i.p}}$  from  $\mathbb{R}^m$  to itself. Specifically, set

$$\mathbf{G}_{\text{s.i.p}}(\mathbf{c}) := \left[ \left[ g_j^\sharp, \sum_{k \in \mathbb{N}_m} c_k g_k^\sharp \right]_{\mathcal{B}^*} : j \in \mathbb{N}_m \right], \quad \text{for all } \mathbf{c} := [c_k : k \in \mathbb{N}_m] \in \mathbb{R}^m. \quad (42)$$

We present the solution methods for this case in the following two results.

**Corollary 59** Suppose that  $\mathcal{B}$  is a uniformly Fréchet smooth and uniformly convex Banach space and  $g_j \in \mathcal{B}$ ,  $j \in \mathbb{N}_m$ . Let  $\mathcal{L}$  be the linear operator defined by (1) with  $\nu_j := g_j^\sharp$ ,  $j \in \mathbb{N}_m$ ,  $\mathcal{L}^*$  be the adjoint operator of  $\mathcal{L}$  and  $\mathbf{G}_{\text{s.i.p}}$  be the operator defined by (42). For a given  $\mathbf{y}_0 \in \mathbb{R}^m$ , let  $\mathcal{Q}_{\mathbf{y}_0} : \mathbb{R}^m \rightarrow \mathbb{R}_+$  be a loss function,  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a strictly increasing

regularizer and  $\lambda > 0$ . Then  $f_0 := \left( \sum_{j \in \mathbb{N}_m} \hat{c}_j g_j^\# \right)^\#$ , is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  if and only if  $\hat{\mathbf{c}} := [\hat{c}_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$  is a solution of the minimization problem

$$\inf \left\{ \mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}_{\text{s.i.p.}}(\mathbf{c})) + \lambda \varphi \left( (\mathbf{c}^\top \mathbf{G}_{\text{s.i.p.}}(\mathbf{c}))^{1/2} \right) : \mathbf{c} \in \mathbb{R}^m \right\}. \quad (43)$$

In the following, we show that the finite dimensional minimization problem (43) reduces to a nonlinear system in a special case that both  $\mathcal{Q}_{\mathbf{y}_0}$  and  $\varphi$  are convex and differentiable.

**Remark 60** For a given  $\mathbf{y}_0 \in \mathbb{R}^m$ , let  $\mathcal{Q}_{\mathbf{y}_0}$  be a convex loss function and  $\varphi$  be a strictly increasing and convex regularizer. If  $\mathcal{Q}_{\mathbf{y}_0}$  and  $\varphi$  are both differentiable, then  $\hat{\mathbf{c}} \neq 0$  is the unique solution of the minimization problem (43) if and only if  $\hat{\mathbf{c}}$  is the solution of the nonlinear system

$$\nabla \mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}_{\text{s.i.p.}}(\hat{\mathbf{c}})) + \lambda \frac{\varphi' \left( (\hat{\mathbf{c}}^\top \mathbf{G}_{\text{s.i.p.}}(\hat{\mathbf{c}}))^{1/2} \right)}{(\hat{\mathbf{c}}^\top \mathbf{G}_{\text{s.i.p.}}(\hat{\mathbf{c}}))^{1/2}} \hat{\mathbf{c}} = 0. \quad (44)$$

The nonlinear system was established in Zhang and Zhang (2012) in the case that  $\mathcal{B}$  is a semi-inner-product RKBS and for  $\mathbf{y}_0 := [y_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$ , the loss function  $\mathcal{Q}_{\mathbf{y}_0}$  has the form  $\mathcal{Q}_{\mathbf{y}_0}(\mathbf{z}) := \sum_{j \in \mathbb{N}_m} \mathcal{Q}_j(z_j, y_i)$ , for all  $\mathbf{z} := [z_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$ , where  $\mathcal{Q}_j : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$ ,  $j \in \mathbb{N}_m$ , are a finite number of bivariate loss functions.

Below, we develop a fixed-point approach for solving the regularization problem in a Banach space. Following the idea in section 5, we now consider solving directly the regularization problem (35) in a Banach space  $\mathcal{B}$ . We will characterize the solutions of the problem via fixed-point equations. Again, we need to consider both cases when the loss function  $\mathcal{Q}_{\mathbf{y}_0}$  is differentiable and when it is not differentiable.

**Theorem 61** Suppose that  $\mathcal{B}$  is a Banach space with the dual space  $\mathcal{B}^*$ ,  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$  and that  $\mathcal{L}$  is defined by (1),  $\mathcal{L}^*$  is the adjoint operator of  $\mathcal{L}$  and  $\mathcal{V}$  is defined earlier. Let  $\mathcal{H}$  be a Hilbert space and  $\mathcal{T}$  a bounded linear operator from  $\mathcal{B}$  to  $\mathcal{H}$  such that  $\mathcal{T}^* \mathcal{T}$  is a one-to-one mapping from the inverse image of  $\mathcal{V}$  onto  $\mathcal{V}$ . For a given  $\mathbf{y}_0 \in \mathbb{R}^m$ , let  $\mathcal{Q}_{\mathbf{y}_0} : \mathbb{R}^m \rightarrow \mathbb{R}_+$  be a convex loss function,  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$  be a convex regularizer and  $\lambda > 0$ . Then  $f_0 \in \mathcal{B}$  is a solution of problem (35) with  $\mathbf{y} := \mathbf{y}_0$  if and only if there exists  $\hat{\mathbf{c}} \in \mathbb{R}^m$  such that

$$\hat{\mathbf{c}} = \text{prox}_{\mathcal{Q}_{\mathbf{y}_0}^*}(\hat{\mathbf{c}} + \mathcal{L}(f_0)) \quad \text{and} \quad f_0 = \text{prox}_{\varphi \circ \|\cdot\|_{\mathcal{B}, \mathcal{H}, \mathcal{T}}} \left( f_0 - \frac{1}{\lambda} (\mathcal{T}^* \mathcal{T})^{-1} \mathcal{L}^*(\hat{\mathbf{c}}) \right). \quad (45)$$

In the special case when the loss function  $\mathcal{Q}_{\mathbf{y}_0}$  is differentiable, the fixed-point equations (45) can reduce to only one fixed-point equation.

**Corollary 62** Suppose that the hypotheses of Theorem 61 hold. If in addition the loss function  $\mathcal{Q}_{\mathbf{y}_0} : \mathbb{R}^m \rightarrow \mathbb{R}_+$  is differentiable, then  $f_0 \in \mathcal{B}$  is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  if and only if it satisfies the fixed-point equation

$$f_0 = \text{prox}_{\varphi \circ \|\cdot\|_{\mathcal{B}, \mathcal{H}, \mathcal{T}}} \left( f_0 - \frac{1}{\lambda} (\mathcal{T}^* \mathcal{T})^{-1} \mathcal{L}^* \nabla \mathcal{Q}_{\mathbf{y}_0} \mathcal{L}(f_0) \right). \quad (46)$$

In particular, for the learning network problem, in which the loss function  $\mathcal{Q}_{\mathbf{y}_0}$  has the form (36), we have the following special result. Note that in this case there holds  $\nabla \mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(f_0)) = 2(\mathcal{L}(f_0) - \mathbf{y}_0)$ .

**Remark 63** *If for a given  $\mathbf{y}_0 \in \mathbb{R}^m$ , the loss function  $\mathcal{Q}_{\mathbf{y}_0}$  has the form (36), the fixed-point equation (46) reduces to*

$$f_0 = \text{prox}_{\varphi_0 \|\cdot\|_{\mathcal{B}, \mathcal{H}, \mathcal{T}}} \left( f_0 - \frac{2}{\lambda} (\mathcal{T}^* \mathcal{T})^{-1} \mathcal{L}^*(\mathcal{L}(f_0) - \mathbf{y}_0) \right).$$

According to Theorem 61 the solution of the regularization problem (35) can be obtained by solving fixed-point equations (45) or (46). Note that either the second equation in (45) or equation (46) is of infinite dimension. In section 5, we have demonstrated that a solution of problem (3) with  $\mathcal{B} = \ell_1(\mathbb{N})$  can be formulated as finite dimensional fixed-point equations. We next show that a solution of the regularization problem (35) with  $\mathcal{B} = \ell_1(\mathbb{N})$  can also be formulated as finite dimensional fixed-point equations. The regularization problem in the space  $\ell_1(\mathbb{N})$  has the form

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

With the help of Lemma 74 stated in Appendix E, a solution of the regularization problem (47) can be characterized via finite dimensional fixed-point equations as follows.

**Theorem 64** *Suppose that  $\mathbf{u}_j \in c_0$ ,  $j \in \mathbb{N}_m$ ,  $\mathcal{L}$  is defined by (14) and  $\mathcal{L}^*$  is the adjoint operator. Let  $\mathcal{T}_0$  be defined by (28) and  $\mathcal{S}$  be the truncation operator. For a given  $\mathbf{y}_0 \in \mathbb{R}^m$ , let  $\mathcal{Q}_{\mathbf{y}_0} : \mathbb{R}^m \rightarrow \mathbb{R}_+$  be a convex loss function and  $\lambda > 0$ . Then  $\mathbf{x}_0 \in \ell_1(\mathbb{N})$  is a solution of the regularization problem (47) with  $\mathbf{y} := \mathbf{y}_0$  if and only if there exists  $\hat{\mathbf{c}} \in \mathbb{R}^m$  such that*

$$\hat{\mathbf{c}} = \text{prox}_{\mathcal{Q}_{\mathbf{y}_0}^*}(\hat{\mathbf{c}} + \mathcal{L}(\mathbf{x}_0)) \quad \text{and} \quad \mathbf{x}_0 = \text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0} \left( \mathbf{x}_0 - \frac{1}{\lambda} \mathcal{S} \mathcal{L}^*(\hat{\mathbf{c}}) \right). \quad (48)$$

If the loss function  $\mathcal{Q}_{\mathbf{y}_0}$  is differentiable, the solution of the regularization problem (47) with  $\mathbf{y} := \mathbf{y}_0$  can be characterized via a single fixed-point equation. We present this result in the next corollary.

**Corollary 65** *Suppose that the hypotheses of Theorem 64 hold. If in addition the loss function  $\mathcal{Q}_{\mathbf{y}_0} : \mathbb{R}^m \rightarrow \mathbb{R}_+$  is differentiable, then  $\mathbf{x}_0 \in \ell_1(\mathbb{N})$  is a solution of the regularization problem (47) with  $\mathbf{y} := \mathbf{y}_0$  if and only if*

$$\mathbf{x}_0 = \text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0} \left( \mathbf{x}_0 - \frac{1}{\lambda} \mathcal{S} \mathcal{L}^* \nabla \mathcal{Q}_{\mathbf{y}_0} \mathcal{L}(\mathbf{x}_0) \right). \quad (49)$$

Once again, for the learning network problem, in which the loss function  $\mathcal{Q}_{\mathbf{y}_0}$  has the form (36), we have the following special result.

**Remark 66** *If for a given  $\mathbf{y}_0 \in \mathbb{R}^m$ , the loss function  $\mathcal{Q}_{\mathbf{y}_0}$  has the form (36), the fixed-point equation (49) reduces to*

$$\mathbf{x}_0 = \text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0} \left( \mathbf{x}_0 - \frac{2}{\lambda} \mathcal{S} \mathcal{L}^*(\mathcal{L}(\mathbf{x}_0) - \mathbf{y}_0) \right). \quad (50)$$

Below, we point out the finite dimensional component of the fixed-points equations (48), the same as those for MNI stated in Theorem 38. To see this, we rewrite the fixed-point equations (48) in the following compact form  $\mathbf{s}_r = (\mathcal{P}_r \circ \mathcal{R}_r)(\mathbf{s}_r)$ , where  $\mathbf{s}_r$  denotes the vector  $\mathbf{s}_r := \begin{bmatrix} \hat{\mathbf{c}} \\ \mathbf{x}_0 \end{bmatrix}$  and two matrices  $\mathcal{P}_r$  and  $\mathcal{R}_r$  of operators have the form

$$\mathcal{P}_r := \begin{bmatrix} \text{prox}_{\mathcal{Q}_{\mathbf{y}_0}^*} \\ \text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0} \end{bmatrix} \quad \text{and} \quad \mathcal{R}_r := \begin{bmatrix} \mathcal{I} & \mathcal{L} \\ -\frac{1}{\lambda} \mathcal{S} \mathcal{L}^* & \mathcal{I} \end{bmatrix}. \quad (51)$$

We show in the following theorem that the compact form fixed-point equation is of finite dimension. This theorem may be proved by similar arguments in the proof of Theorem 38 and we omit the details.

**Theorem 67** *If operators  $\mathcal{P}_r$  and  $\mathcal{R}_r$  are defined as in (51), then  $\mathcal{P}_r \circ \mathcal{R}_r$  is an operator from  $(\mathbb{R}^m, \ell_1(\mathbb{N}))$  to  $(\mathbb{R}^m, c_c)$  and its fixed-point  $\mathbf{s}_r = \begin{bmatrix} \hat{\mathbf{c}} \\ \mathbf{x}_0 \end{bmatrix} \in (\mathbb{R}^m, \ell_1(\mathbb{N}))$  satisfies  $\mathbf{x}_0 \in c_c$  and  $\text{supp}(\mathbf{x}_0) \subseteq \text{supp}(\mathcal{S}(\mathcal{L}^*(\hat{\mathbf{c}})))$ .*

A solution of the regularization problem (47) with  $\mathcal{B} := \ell_1(\mathbb{N})$  guaranteed by Theorems 64 and 67 has an additional property.

**Remark 68** *Each solution  $\mathbf{x}_0 \in \ell_1(\mathbb{N})$  of the regularization problem (47) together with  $\hat{\mathbf{c}} \in \mathbb{R}^m$  satisfying the fixed-point equations (48) is of finite dimension, that is, it satisfies  $\mathbf{x}_0 \in c_c$  and  $\text{supp}(\mathbf{x}_0) \subseteq \text{supp}(\mathcal{S}(\mathcal{L}^*(\hat{\mathbf{c}})))$ .*

Theorem 64 provides a theoretical foundation for algorithmic development for solving the regularization problem (35) with  $\mathcal{B} := \ell_1(\mathbb{N})$ . Specifically, the fixed-point equations (48) on the finite dimensional space will serve as a starting point to design efficient fixed-point iterative algorithms. We postpone further algorithmic development for a future project.

To close this section, we comment on closed-form formulas for the proximity operator of loss functions, required to find a fixed-point according to equations (48). The closed-form of  $\text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0}$  has been given in (30). When the loss function is not differentiable, we also need a closed-form formula for its proximity operator. The proximity operator of certain commonly used loss functions can also be computed explicitly (Li et al., 2019). For example, if  $\mathcal{Q}_{\mathbf{y}}$  is defined as in (38), the proximity operator  $\text{prox}_{\mathcal{Q}_{\mathbf{y}}}$  at  $\mathbf{a} := [a_j : j \in \mathbb{N}_m]$  has the form  $\text{prox}_{\mathcal{Q}_{\mathbf{y}}}(\mathbf{a}) := [b_j : j \in \mathbb{N}_m]$ , where for all  $j \in \mathbb{N}_m$ ,  $b_j := a_j + y_j$ , if  $y_j a_j < 1 - y_j^2$ ,  $b_j := 1/y_j$ , if  $1 - y_j^2 \leq y_j a_j < 1$ ,  $b_j := a_j$ , if  $y_j a_j \geq 1$ . If  $\mathcal{Q}_{\mathbf{y}}$  is defined as in (37), we present the proximity operator  $\text{prox}_{\mathcal{Q}_{\mathbf{y}}}$  at  $\mathbf{a} := [a_j : j \in \mathbb{N}_m]$  as follows. If  $\epsilon \geq 1/2$ , then for all  $j \in \mathbb{N}_m$

$$b_j := \begin{cases} a_j + 1, & \text{if } a_j < -\epsilon - 1 + y_j, \\ -\epsilon + y_j, & \text{if } -\epsilon - 1 + y_j \leq a_j < -\epsilon + y_j, \\ a_j + 1, & \text{if } -\epsilon + y_j \leq a_j < \epsilon - 1 + y_j, \\ a_j, & \text{if } \epsilon - 1 + y_j \leq a_j < \epsilon + y_j, \\ \epsilon + y_j, & \text{if } \epsilon + y_j \leq a_j < \epsilon + 1 + y_j, \\ a_j - 1, & \text{if } a_j \geq \epsilon + 1 + y_j. \end{cases}$$

If  $\epsilon < 1/2$ , then for all  $j \in \mathbb{N}_m$

$$b_j := \begin{cases} a_j + 1, & \text{if } a_j < -\epsilon - 1 + y_j, \\ -\epsilon + y_j, & \text{if } -\epsilon - 1 + y_j \leq a_j < -\epsilon + y_j, \\ a_j, & \text{if } -\epsilon + y_j \leq a_j < \epsilon + y_j, \\ \epsilon + y_j, & \text{if } \epsilon + y_j \leq a_j < \epsilon + 1 + y_j, \\ a_j - 1, & \text{if } a_j \geq \epsilon + 1 + y_j. \end{cases}$$

## 8. Note on the Related Existing Work

In this note, we identify the relation of the representer theorems obtained in this paper with those that have already existed in the literature.

Certain forms of implicit solution representations of the MNI problem in some Banach spaces exist in the literature. In Proposition 7 that summarizes four implicit solution representations of the MNI problem in Banach spaces, the equivalence of (i) and (iii) was established in Micchelli and Pontil (2004). From Remark 11 which gives a special explicit representer theorem when  $\mathcal{B}$  is a Banach space having the smooth dual space  $\mathcal{B}^*$ , one can immediately obtain the classical representer theorem for the MNI problem in an RKHS  $\mathcal{H}$  (Wendland, 2005), by employing the facts that  $\mathcal{H} = \mathcal{H}^*$  and  $f = \|f\|_{\mathcal{H}}\mathcal{G}(f)$  for any  $f \in \mathcal{H}$  in this special case. The implicit representer theorems for the MNI problem in a semi-inner-product RKBS, stated in (1) of Corollary 16 was originally obtained in Zhang et al. (2009) by a different approach: the orthogonality in a semi-inner-product RKBS, characterized through the dual element and the semi-inner-product. The implicit representer theorems for the MNI problem in a right-sided RKBS, stated in Corollary 17, were originally obtained in Xu and Ye (2019) by a different approach: the orthogonality in these Banach spaces, described through the Gâteaux derivatives and reproducing properties.

Likewise, certain forms of implicit solution characterizations for the regularization problem have also been established in some Banach spaces, in the literature. Among the four implicit solution characterizations established in Proposition 42 for the regularization problem in a general Banach space, (iv) for a special regularizer  $\varphi(t) := t^2$ ,  $t \in \mathbb{R}_+$ , was obtained in Huang et al. (2020) and (iii) was derived in Unser (2019b) via a different approach, the duality mapping. The well-known representer theorem in an RKHS which was originally established in Kimeldorf and Wahba (1970) and was generalized for non-quadratic loss functions and nondecreasing regularizers (Argyriou et al., 2009; Cox and O’Sullivan, 1990; Schölkopf et al., 2001) can be obtained from Remark 45 by specifying the smooth Banach space  $\mathcal{B}$  to be the RKHS  $\mathcal{H}$  and noting that  $\mathcal{H} = \mathcal{H}^*$  and  $f = \|f\|_{\mathcal{H}}\mathcal{G}(f)$  for any  $f \in \mathcal{H}$  in this special case. Remark 45 also leads directly to the representer theorem for regularization problems in an FRKHS from a finite number of non-point-evaluation functional data, which was originally established in Wang and Xu (2019). The implicit representer theorem for the regularization problem in a semi-inner-product RKBS (stated in Corollary 50) was originally proved in Zhang et al. (2009) and Zhang and Zhang (2012) and that in a right-sided RKBS (stated in Corollary 52) was established in Xu and Ye (2019), both by a different approach: the orthogonality in Banach spaces.

The explicit representer theorems for Banach spaces have not been seen in the literature, to our best knowledge, except for the Hilbert spaces.

## Acknowledgments

R. Wang is supported by National Key R&D Program of China (No. 2020YFA0714101), by Natural Science Foundation of China under grant 12171202, by the Fundamental Research Funds for the Central Universities and by the Opening Project of Guangdong Province Key Laboratory of Computational Science at the Sun Yat-sen University under grant 2018005. Y. Xu is supported by US National Science Foundation under grant DMS-1912958 and by Natural Science Foundation of China under grant 11771464. Y. Xu is the correspondence author.

## Appendix A. Sparse Classification in Reproducing Kernel Banach Spaces

We discuss briefly in this appendix the relevance of RKBSs to machine learning, by recalling an example of sparse classification.

We consider a typical machine learning problem of classifying two groups of data points. Suppose that training data  $D := \{(x_k, y_k) : k \in \mathbb{N}_n\}$  composed of input data points  $X := \{x_k : k \in \mathbb{N}_n\} \subset \mathbb{R}^d$  and output data values  $Y := \{y_k : k \in \mathbb{N}_n\} \subset \{-1, 1\}$  are given. We wish to find a hyperplane  $s(x) = 0$ , where

$$s(x) := w_s^T x - b_s, \quad \text{for all } x \in \mathbb{R}^d, \quad (52)$$

that separates the data  $D$  into two groups: one with label  $y_k = 1$  which corresponds to  $s(x_k) > 0$  and the other with label  $y_k = -1$  which corresponds to  $s(x_k) < 0$ .

We first review an RKHS approach to this problem. The parameters  $(w_s, b_s) \in \mathbb{R}^d \times \mathbb{R}$  appearing in function  $s$  defined by (52) are chosen such that the hyperplane maximizes its distances to the two groups of points in  $D$ . The function  $s$  will give us a decision rule

$$r(x) := \text{sign}(s(x)), \quad \text{for } x \in \mathbb{R}^d \quad (53)$$

to predict labels for new data points. To determine the parameters  $(w_s, b_s) \in \mathbb{R}^d \times \mathbb{R}$ , we select two hyperplanes  $w^T x - b = 1$  and  $w^T x - b = -1$  that both parallel to  $s(x) = 0$  and separate the two groups of data in the way that the distance between them is as large as possible. The region bounded by these two hyperplanes is called the margin. Note that the distance between these two hyperplanes is  $\frac{2}{\|w\|_2}$ . Maximizing the distance between the two hyperplanes is equivalent to minimizing  $\frac{\|w\|_2}{2}$ . To prevent data points from falling into the margin, we add the constraints

$$y_k(w^T x_k - b) \geq 1, \quad k \in \mathbb{N}_n. \quad (54)$$

Thus, the parameters  $(w_s, b_s)$  are obtained by solving a constrained minimization problem

$$\min \left\{ \frac{1}{2} \|w\|_2 : (w, b) \in \mathbb{R}^d \times \mathbb{R}, \text{ subject to constraints (54)} \right\}. \quad (55)$$

The constrained minimization problem (55) may be reformulated in a regularization form by using the hinge loss function

$$L(y, t) := \max\{0, 1 - yt\}, \quad \text{for } y \in \{-1, 1\} \text{ and } t \in \mathbb{R},$$

to write the constraints (54) as a fidelity term. That is, we consider regularization problem

$$(w_s, b_s) := \operatorname{argmin} \left\{ \frac{1}{n} \sum_{k \in \mathbb{N}_n} L(y_k, w^T x_k - b) + \lambda \|w\|_2^2 : (w, b) \in \mathbb{R}^d \times \mathbb{R} \right\}. \quad (56)$$

The solution  $(w_s, b_s)$  of (56) yields the function  $s$  for the decision rule (53).

Model (56) may lead to misclassification, since not all data sets in  $\mathbb{R}^d$  can be separated by a hyperplane in the same space  $\mathbb{R}^d$ . To alleviate misclassification, one possible approach is to map the given data sets in  $\mathbb{R}^d$  to a higher dimensional space (or even an infinite dimensional space) and perform classification in the new space. The idea is to choose an appropriate feature map  $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^D$  with  $d \ll D$ . When  $D = +\infty$ ,  $\mathbb{R}^D$  will be replaced by a Hilbert space  $\mathcal{H}$ . We shall discuss the choice of the feature map below.

It is worth to comment the advantage of developing learning methods in an infinite-dimensional space of functions over a finite-dimensional space. An infinite-dimensional space of functions has much more capacity of representing a function than a finite dimensional one. In particular, for the classification problem, given data sets are more likely separable by a hyperplane in an infinite dimensional space than by that in a finite dimensional one. One may argue that a learning method in an infinite dimensional space may require more computational complexity to implement it than in a finite dimensional one. This may be the case if the infinite dimensional space is arbitrarily chosen. When an RKHS is used, due to the masterful representer theorem of learning in an RKHS, the learning method in such a space reduces to finding a finite number of coefficients of elements in it. It is the main focus of this article to study representer theorems of learning in RKBSs.

Let us return to our original problem. Due to the fact that function values are used, we would require that the Hilbert space of functions that we choose to work with has the property that the point evaluation functionals are continuous in the space. This naturally leads to the choice of a RKHS that is a Hilbert space in which the point evaluation functionals are continuous. For this reason, we choose a RKHS  $\mathcal{H}$ , with the reproducing kernel  $K$ , as the feature space and  $\Phi : \mathbb{R}^d \rightarrow \mathcal{H}$ , defined by  $\Phi(x) := K(x, \cdot)$ , as the feature map. In this way, one can extend the regularization problem (56) to

$$\min \left\{ \frac{1}{n} \sum_{k \in \mathbb{N}_n} L(y_k, f(x_k)) + \lambda \|f\|_{\mathcal{H}}^2 : f \in \mathcal{H} \right\} \quad (57)$$

to determine a function  $f$  that define a decision rule. The representer theorem (Schölkopf et al., 2001) yields the representation of the optimal solution  $f^*$  of the optimization problem (57) as

$$f^*(x) = \sum_{k \in \mathbb{N}_n} c_k K(x_k, x), \quad (58)$$

for some suitable parameters  $c_k \in \mathbb{R}$ ,  $k \in \mathbb{N}_n$ . We remark that the regularization problem (56) with  $b = 0$  is a special case of (57), with  $K(x, y) := \langle x, y \rangle$ , for all  $x, y \in \mathbb{R}^d$ , and  $\mathcal{H} := \{\langle \cdot, x \rangle : x \in \mathbb{R}^d\}$ ,  $f(x) = w^T x$ , for  $x \in \mathbb{R}^d$ , and  $\|f\|_{\mathcal{H}} = \|w\|_2$ .

We next describe how the RKBS comes into play in classification. Learning methods in an RKHS result in dense representations (58) of learning solutions, that is, most of the

terms in (58) are nonzero. A dense representation of a learning solution requires large computational costs to implement it. In the context of big data analytics, it is desirable to have sparse learning methods. To this end, we appeal to Banach spaces, since nonsmoothness of certain Banach spaces such as the 1-norm space may lead to sparse learning methods. This may be illustrated by an example given in (Cheng and Xu, 2021). Specifically, consider seeking  $x_p^* \in \ell_p(\mathbb{N})$  for  $p = 1, 2$  such that

$$\|x_p^*\|_p = \inf\{\|x\|_p : x \in \ell_p(\mathbb{N}), \langle u_j, x \rangle_{\ell_p} = y_j, j = 1, 2\}, \quad (59)$$

where  $y_1 = 3$ ,  $y_2 = 4$ ,  $u_1 = (\frac{1}{n} : n \in \mathbb{N})$ ,  $u_2 = (\frac{1}{(-2)^{n-1}} : n \in \mathbb{N})$ . The solutions of problem (59) with  $p = 1$  and  $p = 2$  are given by  $x_1^* = (\frac{7}{2}, -1, 0, 0, \dots)$  and  $x_2^* = (\frac{0.4924584}{n} + \frac{2.7004714}{(-2)^{n-1}} : n \in \mathbb{N})$ , respectively. Clearly, the solution  $x_1^*$  in the Banach space  $\ell_1(\mathbb{N})$  is sparse while the solution  $x_2^*$  in the Hilbert space  $\ell_2(\mathbb{N})$  is dense. Solutions of regularized learning methods in the 1-norm space also exhibit sparsity but those in the 2-norm space do not.

Aiming at obtaining a sparse solution, instead of solving the regularization problem (57), we consider solving the regularization problem

$$\min \left\{ \frac{1}{n} \sum_{k \in \mathbb{N}_n} L(y_k, f(x_k)) + \lambda \|f\|_{\mathcal{B}} : f \in \mathcal{B} \right\}, \quad (60)$$

where  $\mathcal{B}$  denotes an RKBS. Problems (57) and (60) have the same fidelity term but different regularization terms, one in an RKHS and one in an RKBS. The solution of the regularization problem (60) may lead to sparse classification solutions if  $\mathcal{B}$  is chosen to be the 1-norm RKBS. The reason for which a reproducing kernel is required is due to the use of the function values  $f(x_k)$  in the fidelity term of problem (60). Specific numerical examples of classification in RKBSs which demonstrate the accuracy and sparsity of the resulting learning solutions may be found in (Li et al., 2018; Xu and Ye, 2019; Lin et al., 2021).

To close this section, we mention a recent paper Adcock et al. (2021) which studied infinite-dimensional compressed sensing. According to Adcock et al. (2021), there are advantages of using an infinite-dimensional compressed sensing model for analog problems. Although much of the compressive imaging literature considers the recovery of discrete images from discrete measurements, modalities such as MRI and NMR are naturally analog. Hence, better modeled them over the continuum. Applying finite-dimensional recovery procedures to analog problems can result in artefacts. The setting of MNI or regularization problems in Banach spaces considered in this paper covers the infinite-dimensional compressed sensing model.

## Appendix B. Proofs for Section 2

**Proof** [of Proposition 1] We prove the existence for the case that  $\mathcal{B}$  has a separable pre-dual space  $\mathcal{B}_*$  by a weak\* minimizing sequence argument. Since for any  $\mathbf{y} \in \mathbb{R}^m$ , the set  $\mathcal{M}_{\mathbf{y}}$  is nonempty, there exists a sequence  $f_n$ ,  $n \in \mathbb{N}$ , in  $\mathcal{M}_{\mathbf{y}}$  satisfying  $\lim_{n \rightarrow +\infty} \|f_n\|_{\mathcal{B}} = \inf\{\|f\|_{\mathcal{B}} : f \in \mathcal{M}_{\mathbf{y}}\}$ . This ensures that the sequence  $f_n$ ,  $n \in \mathbb{N}$ , is bounded. By the Banach-Alaoglu theorem, there exists a subsequence  $f_{n_k}$ ,  $k \in \mathbb{N}$ , which weakly\* converges to  $\hat{f} \in \mathcal{B}$ . It suffices to prove that the weak\* accumulation point  $\hat{f}$  is a solution of the MNI problem (3).



We first verify that  $\hat{f}$  satisfies the interpolation condition. Since  $\nu_j \in \mathcal{B}_*$ ,  $j \in \mathbb{N}_m$ , the linear functionals  $\nu_j$  are weakly\* continuous. This leads to  $\langle \nu_j, \hat{f} \rangle_{\mathcal{B}} = \lim_{k \rightarrow +\infty} \langle \nu_j, f_{n_k} \rangle_{\mathcal{B}}$ , for all  $j \in \mathbb{N}_m$ . By the fact that  $f_n \in \mathcal{M}_{\mathbf{y}}$ ,  $n \in \mathbb{N}$ , we get  $\mathcal{L}(\hat{f}) = \mathbf{y}$ . That is, the interpolation condition holds. Note that the norm  $\|\cdot\|_{\mathcal{B}}$  is weakly\* lower semi-continuous on  $\mathcal{B}$ . According to the weak\* convergence of  $f_{n_k}$ ,  $k \in \mathbb{N}$ , there holds  $\|\hat{f}\|_{\mathcal{B}} \leq \liminf_{k \rightarrow +\infty} \|f_{n_k}\|_{\mathcal{B}}$ , which yields that  $\hat{f}$  is a solution of problem (3).

In the case that the pre-dual space  $\mathcal{B}_*$  is non-separable, the desired result may be proved by the generalized Weierstrass theorem in the weak\* topology (Kurdila and Zabaranin, 2005).  $\blacksquare$

### Appendix C. Proofs for Section 3

**Proof** [of Proposition 2] We first remark that the relation between  $\mathcal{M}_{\mathbf{y}}$  and  $\mathcal{M}_0$  leads to the following fact: If  $\mathbf{y} \in \mathbb{R}^m \setminus \{0\}$ , then  $d(f, \mathcal{M}_0) = d(g, \mathcal{M}_{\mathbf{y}})$ , for all  $f \in \mathcal{M}_{\mathbf{y}}$  and  $g \in \mathcal{M}_0$ .

An element  $\hat{f} \in \mathcal{B}$  is a solution of problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and  $\|\hat{f}\|_{\mathcal{B}} = \inf\{\|f\|_{\mathcal{B}} : f \in \mathcal{M}_{\mathbf{y}}\} = d(0, \mathcal{M}_{\mathbf{y}})$ . By the fact established earlier, we get that  $\hat{f}$  is a solution of (3) if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and  $\|\hat{f}\|_{\mathcal{B}} = d(\hat{f}, \mathcal{M}_0)$ . The latter is equivalent to  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and 0 being a best approximation to  $\hat{f}$  from  $\mathcal{M}_0$ .  $\blacksquare$

To prove Lemma 4, we recall Proposition 2.6.6 of Megginson (1998) which states that for a subset  $\mathcal{M}'$  in  $\mathcal{B}^*$ , the set  $(\mathcal{M}')^{\perp}$  coincides with the closed linear span of  $\mathcal{M}'$  in the weak\* topology of  $\mathcal{B}^*$ . For a subset  $\mathcal{M}'$  of  $\mathcal{B}^*$ , we denote by  $\overline{\mathcal{M}'}^{w*}$  the closure of  $\mathcal{M}'$  in the weak\* topology of  $\mathcal{B}^*$ .

**Proof** [of Lemma 4] We prove this lemma by appealing to Proposition 2.6.6 of Megginson (1998). The definition of the annihilator leads to  $\mathcal{M}_0 = \mathcal{V}_m^{\perp}$ . Applying Proposition 2.6.6 of Megginson (1998) with  $\mathcal{M}' := \mathcal{V}_m$  yields that  $\mathcal{M}_0^{\perp} = (\mathcal{V}_m^{\perp})^{\perp} = \overline{\text{span } \mathcal{V}_m}^{w*}$ . Since the linear span of  $\mathcal{V}_m$  is a finite dimensional subspace of  $\mathcal{B}^*$ , there holds  $\overline{\text{span } \mathcal{V}_m}^{w*} = \text{span } \mathcal{V}_m$ . Substituting this equation into the right hand side in the equation above leads to the desired result of this lemma.  $\blacksquare$

**Proof** [of Proposition 5] We first consider the case that  $\mathbf{y} := [y_j : j \in \mathbb{N}_m] = 0$ . Note that the MNI problem (3) with  $\mathbf{y} = 0$  has a unique solution  $\hat{f} = 0$ . On one hand, it is clear that the trivial solution  $\hat{f} = 0$  belongs to  $\mathcal{M}_0$ . Moreover, equations (6) also hold by choosing  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$  such that the norm of the functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  equals to 1. On the other hand, if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that the functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  satisfying  $\langle \nu, \hat{f} \rangle_{\mathcal{B}} = \|\hat{f}\|_{\mathcal{B}}$ , then  $\|\hat{f}\|_{\mathcal{B}} = \sum_{j \in \mathbb{N}_m} c_j \langle \nu_j, \hat{f} \rangle_{\mathcal{B}} = \sum_{j \in \mathbb{N}_m} c_j y_j = 0$ , which further implies  $\hat{f} = 0$ . That is, we get the desired conclusion for  $\mathbf{y} = 0$ .

If  $\mathbf{y} \neq 0$ , Proposition 3 ensures that  $\hat{f} \in \mathcal{B}$  is a solution of the MNI problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there is a continuous linear functional  $\nu \in \mathcal{M}_0^{\perp}$  satisfying equations (6). By Lemma 4, there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that this continuous linear functional  $\nu$  has the form  $\nu = \sum_{j \in \mathbb{N}_m} c_j \nu_j$ . This establishes the desired result of this proposition.  $\blacksquare$

To prove Theorem 6, we review the Lagrange multiplier method as follows.

**Lemma 69** *If  $\phi$  and  $\psi_j$ ,  $j \in \mathbb{N}_m$ , are all convex functions from  $\mathcal{B}$  to  $\mathbb{R}$ , then  $\hat{f} \in \mathcal{B}$  is a solution of the optimization problem*

$$\inf\{\phi(f) : f \in \mathcal{B}, \psi_j(f) = 0, j \in \mathbb{N}_m\}$$

*if and only if  $\psi_j(\hat{f}) = 0$ ,  $j \in \mathbb{N}_m$ , and there exist  $\lambda_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $0 \in \partial\left(\phi + \sum_{j \in \mathbb{N}_m} \lambda_j \psi_j\right)(\hat{f})$ . Moreover, if  $\psi_j$ ,  $j \in \mathbb{N}_m$ , are continuous at  $\hat{f}$ , then the inclusion is equivalent to  $0 \in \partial\phi(\hat{f}) + \sum_{j \in \mathbb{N}_m} \lambda_j \partial\psi_j(\hat{f})$ .*

By Lemma 69 we provide a complete proof of Theorem 6 in the following.

**Proof** [of Theorem 6] According to Lemma 69 with  $\phi := \|\cdot\|_{\mathcal{B}}$  and  $\psi_j := \langle \nu_j, \cdot \rangle_{\mathcal{B}} - y_j$ ,  $j \in \mathbb{N}_m$ , we have that  $\hat{f}$  is a solution of (3) if and only if  $\langle \nu_j, \hat{f} \rangle_{\mathcal{B}} = y_j$ ,  $j \in \mathbb{N}_m$ , and there exist  $\lambda_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that

$$0 \in \partial\left(\|\cdot\|_{\mathcal{B}} + \sum_{j \in \mathbb{N}_m} \lambda_j (\langle \nu_j, \cdot \rangle_{\mathcal{B}} - y_j)\right)(\hat{f}). \quad (61)$$

Since  $\nu_j$ ,  $j \in \mathbb{N}_m$ , are continuous on  $\mathcal{B}$ , Lemma 69 ensures that equation (61) is equivalent to

$$0 \in \partial\|\cdot\|_{\mathcal{B}}(\hat{f}) + \sum_{j \in \mathbb{N}_m} \lambda_j \partial(\langle \nu_j, \cdot \rangle_{\mathcal{B}} - y_j)(\hat{f}). \quad (62)$$

It follows from the linearity of  $\langle \nu_j, \cdot \rangle_{\mathcal{B}}$ ,  $j \in \mathbb{N}_m$ , that  $\partial\langle \nu_j, \cdot \rangle_{\mathcal{B}}(\hat{f}) = \nu_j$ ,  $j \in \mathbb{N}_m$ . Substituting these equations into (62) leads to  $-\sum_{j \in \mathbb{N}_m} \lambda_j \nu_j \in \partial\|\cdot\|_{\mathcal{B}}(\hat{f})$ . Choosing  $c_j := -\lambda_j$ , for  $j \in \mathbb{N}_m$ , completes the proof of this theorem.  $\blacksquare$

**Proof** [of Proposition 7] The equivalence of statements (i) and (ii) and that of (i) and (v) have been proved in Proposition 5 and Theorem 6, respectively. It remains to verify the equivalence of statements (ii), (iii) and (iv).

We first show the equivalence of statements (ii) and (iii). On one hand, if statement (ii) holds, there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that the linear functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  satisfies the equations in (ii). It is clear that this functional  $\nu$  peaks at  $\hat{f}$ . That is, statement (iii) holds. On the other hand, if statement (iii) holds, there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that the nonzero linear functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  satisfies the equation in (iii). By setting  $\tilde{\mathbf{c}} := \frac{\mathbf{c}}{\|\nu\|_{\mathcal{B}^*}}$  and  $\tilde{\nu} := \sum_{j \in \mathbb{N}_m} \tilde{c}_j \nu_j$ , we get that  $\tilde{\nu}$  satisfies the two equations in (ii) and thus statement (ii) holds.

We next prove the equivalence of statements (ii) and (iv). Note that if  $\hat{f} = 0$ , statements (ii) and (iv) both hold without any assumptions. Specifically, the two equations in (ii) hold by choosing  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that the norm of  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  equals to 1. On the other hand, the inclusion relation in (iv) can be obtained by choosing  $c_j = 0$ ,  $j \in \mathbb{N}_m$ . Hence, it remains to prove the equivalence for the case that  $\hat{f} \neq 0$ . For each  $\mathbf{c} := [c_j : j \in \mathbb{N}_m] \in \mathbb{R}^m$ , we scale it by setting  $\tilde{\mathbf{c}} := \|\hat{f}\|_{\mathcal{B}} \mathbf{c}$ . Set  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  and  $\tilde{\nu} := \sum_{j \in \mathbb{N}_m} \tilde{c}_j \nu_j$ . Clearly,  $\nu$  satisfies the equations in (ii) if and only if  $\tilde{\nu}$  satisfies that  $\|\tilde{\nu}\|_{\mathcal{B}^*} = \|\hat{f}\|_{\mathcal{B}}$  and  $\langle \tilde{\nu}, \hat{f} \rangle_{\mathcal{B}} = \|\tilde{\nu}\|_{\mathcal{B}^*} \|\hat{f}\|_{\mathcal{B}}$ , which is equivalent to that  $\tilde{\nu}$  satisfies the inclusion

relation in (iv). ■

**Proof** [of Theorem 8] If  $\mathbf{y} := [y_j : j \in \mathbb{N}_m] = 0$ , the MNI problem (3) has a unique solution  $\hat{f} = 0$ . It is clear that the trivial solution  $\hat{f} \in \mathcal{M}_0$  and equation  $\mathcal{G}(\hat{f}) = \sum_{j \in \mathbb{N}_m} c_j \nu_j$  holds by choosing  $c_j = 0$ ,  $j \in \mathbb{N}_m$ . Conversely, if  $\hat{f} \in \mathcal{M}_0$  and satisfies  $\mathcal{G}(\hat{f}) = \sum_{j \in \mathbb{N}_m} c_j \nu_j$  for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , we have that  $\langle \mathcal{G}(\hat{f}), \hat{f} \rangle_{\mathcal{B}} = \sum_{j \in \mathbb{N}_m} c_j \langle \nu_j, \hat{f} \rangle_{\mathcal{B}} = \sum_{j \in \mathbb{N}_m} c_j y_j = 0$ , which together with  $\langle \mathcal{G}(\hat{f}), \hat{f} \rangle_{\mathcal{B}} = \|\hat{f}\|_{\mathcal{B}}$  implies  $\hat{f} = 0$ .

We prove this theorem for the case that  $\mathbf{y} \neq 0$  by employing the equivalent conditions (i) and (v) in Proposition 7. To this end, we first show that the subdifferential of the norm  $\|\cdot\|_{\mathcal{B}}$  at any  $f \in \mathcal{B} \setminus \{0\}$  is the singleton  $\mathcal{G}(f)$ , that is,  $\partial\|\cdot\|_{\mathcal{B}}(f) = \{\mathcal{G}(f)\}$ . Suppose that  $\nu \in \partial\|\cdot\|_{\mathcal{B}}(f)$ . Let  $t \in \mathbb{R}$  and  $h \in \mathcal{B}$ . Then the definition of the subdifferentiable of  $\phi := \|\cdot\|_{\mathcal{B}}$  with  $g := f + th$  leads to  $t\langle \nu, h \rangle_{\mathcal{B}} \leq \|f + th\|_{\mathcal{B}} - \|f\|_{\mathcal{B}}$ , which further implies

$$\lim_{t \rightarrow 0^-} \frac{\|f + th\|_{\mathcal{B}} - \|f\|_{\mathcal{B}}}{t} \leq \langle \nu, h \rangle_{\mathcal{B}} \leq \lim_{t \rightarrow 0^+} \frac{\|f + th\|_{\mathcal{B}} - \|f\|_{\mathcal{B}}}{t}.$$

Since  $\mathcal{B}$  is smooth, the norm  $\|\cdot\|_{\mathcal{B}}$  is Gâteaux differentiable at  $f$ . Hence, we get that  $\langle \nu, h \rangle_{\mathcal{B}} = \lim_{t \rightarrow 0} \frac{\|f + th\|_{\mathcal{B}} - \|f\|_{\mathcal{B}}}{t}$ . It follows from equation (9) that  $\nu = \mathcal{G}(f)$ . Due to the arbitrariness of  $\nu \in \partial\|\cdot\|_{\mathcal{B}}(f)$ , we obtain  $\partial\|\cdot\|_{\mathcal{B}}(f) = \{\mathcal{G}(f)\}$ .

According to Proposition 7,  $\hat{f} \in \mathcal{B}$  is a solution of the MNI problem (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that the functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  belongs to  $\partial\|\cdot\|_{\mathcal{B}}(\hat{f})$ . By equation  $\partial\|\cdot\|_{\mathcal{B}}(\hat{f}) = \{\mathcal{G}(\hat{f})\}$  and noting that  $\hat{f} \neq 0$ , the functional  $\nu$  coincides with the Gâteaux derivative of the norm  $\|\cdot\|_{\mathcal{B}}$  at  $\hat{f}$ , which completes the proof of the desired result. ■

**Proof** [of Lemma 9] According to equation (8),  $\nu \in \mathcal{B}^*$  satisfies the inclusion in the left hand side of (10) if and only if  $\left\langle \frac{\nu}{\|\nu\|_{\mathcal{B}^*}}, f \right\rangle_{\mathcal{B}} = \|f\|_{\mathcal{B}}$ , which is equivalent to  $\langle \nu, f \rangle_{\mathcal{B}} = \|\nu\|_{\mathcal{B}^*} \|f\|_{\mathcal{B}}$ . It follows from equation (4) that the above equation is equivalent to  $\langle f, \nu \rangle_{\mathcal{B}^*} = \|\nu\|_{\mathcal{B}^*} \|f\|_{\mathcal{B}}$ , which holds if and only if  $\left\langle \frac{f}{\|f\|_{\mathcal{B}}}, \nu \right\rangle_{\mathcal{B}^*} = \|\nu\|_{\mathcal{B}^*}$ . Again by equation (8) with  $\mathcal{B}$  being replaced by  $\mathcal{B}^*$ , the latter equation is equivalent to that  $f \in \mathcal{B}$  satisfies the inclusion in the right hand side of (10). Consequently, we obtain the equivalence between the two inclusions in (10). ■

**Proof** [of Theorem 10] We first prove this result for  $\mathbf{y} = 0$ . In this case, the MNI problem (3) has a unique solution  $\hat{f} = 0$ . On one hand, if  $\hat{f} = 0$ , there hold  $\hat{f} \in \mathcal{M}_0$  and equation (11) with  $c_j = 0$ ,  $j \in \mathbb{N}_m$ . On the other hand, suppose that  $\hat{f} \in \mathcal{M}_0$  and equation (11) holds for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ . Set  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$ . It follows that  $\langle \nu, \hat{f} \rangle_{\mathcal{B}} = 0$  and  $\hat{f} \in \|\nu\|_{\mathcal{B}^*} \partial\|\cdot\|_{\mathcal{B}}(\nu)$ . If  $\nu = 0$ , the inclusion above leads to  $\hat{f} = 0$ . If  $\nu \neq 0$ , by the inclusion above we have that  $\left\langle \frac{\hat{f}}{\|\hat{f}\|_{\mathcal{B}}}, \nu \right\rangle_{\mathcal{B}^*} = \|\nu\|_{\mathcal{B}^*}$ , which together with (4) leads to  $\langle \nu, \hat{f} \rangle_{\mathcal{B}} = \|\nu\|_{\mathcal{B}^*}^2$ . Since  $\langle \nu, \hat{f} \rangle_{\mathcal{B}} = 0$ , we get that  $\nu = 0$ , which is a contradiction.

We prove this theorem for  $\mathbf{y} \neq 0$  by employing condition (v) in Proposition 7 and Lemma 9. Proposition 7 ensures that  $\hat{f} \in \mathcal{B}$  is a solution of the MNI problem (3) if and

only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $\hat{c}_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that

$$\sum_{j \in \mathbb{N}_m} \hat{c}_j \nu_j \in \partial \|\cdot\|_{\mathcal{B}}(\hat{f}). \quad (63)$$

Set  $c_j := \|\hat{f}\|_{\mathcal{B}} \hat{c}_j$ ,  $j \in \mathbb{N}_m$ . It suffices to prove that  $\hat{c}_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , satisfies (63) if and only if  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , satisfies (11).

On one hand, suppose that  $\hat{c}_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , satisfies (63). Note that since  $\mathbf{y} \neq 0$ ,  $\hat{f} \neq 0$ . Then by equation (63), the functional  $\hat{\nu} := \sum_{j \in \mathbb{N}_m} \hat{c}_j \nu_j$  satisfies that  $\|\hat{\nu}\|_{\mathcal{B}^*} = 1$ . By Lemma 9 we get that

$$\hat{f} \in \|\hat{f}\|_{\mathcal{B}} \partial \|\cdot\|_{\mathcal{B}^*}(\hat{\nu}). \quad (64)$$

Set  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$ . It follows that  $\nu = \|\hat{f}\|_{\mathcal{B}} \hat{\nu}$  which together with  $\|\hat{\nu}\|_{\mathcal{B}^*} = 1$  yields that  $\|\nu\|_{\mathcal{B}^*} = \|\hat{f}\|_{\mathcal{B}}$ . Noting by equation (8) that  $\partial \|\cdot\|_{\mathcal{B}^*}(\hat{\nu}) = \partial \|\cdot\|_{\mathcal{B}^*}(\nu)$ . Substituting this equation and  $\|\nu\|_{\mathcal{B}^*} = \|\hat{f}\|_{\mathcal{B}}$  into (64), we get that  $\hat{f} \in \|\nu\|_{\mathcal{B}^*} \partial \|\cdot\|_{\mathcal{B}^*}(\nu)$ . That is,  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , satisfies (11).

On the other hand, suppose that  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , satisfies (11). Since  $\hat{f} \neq 0$ , we have that the functional  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  are nonzero and then there holds  $\|\nu\|_{\mathcal{B}^*} = \|\hat{f}\|_{\mathcal{B}}$ . By Lemma 9 we obtain that  $\frac{\nu}{\|\nu\|_{\mathcal{B}^*}} \in \partial \|\cdot\|_{\mathcal{B}^*}(\hat{f})$ , which together with the equation above yields that  $\hat{c}_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , satisfies (63).  $\blacksquare$

**Proof** [of Theorem 12] Theorem 10 ensures that  $\hat{f} \in \mathcal{B}$  is a solution of (3) with  $\mathbf{y}$  if and only if  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , satisfying (11). It suffices to show that  $\hat{f} \in \mathcal{B}$  satisfies (11) if and only if it satisfies the desired representation of this theorem. Set  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$ . According to equation (8),  $\hat{f}$  satisfies (11) if and only if  $\langle \hat{f}, \nu \rangle_{\mathcal{B}^*} = \|\nu\|_{\mathcal{B}^*}^2$ . Since  $\nu \in \mathcal{B}_*$ , by equations (4) and (5), the above equation is equivalent to  $\langle \hat{f}, \nu \rangle_{\mathcal{B}^*} = \|\nu\|_{\mathcal{B}_*}^2$ . Again by (8) with  $\mathcal{B}$  being replaced by  $\mathcal{B}_*$ , we get that the above equation holds if and only if  $\hat{f}$  satisfies the desired representation.  $\blacksquare$

**Proof** [of Theorem 14] Since  $\mathcal{B}$  is uniformly convex, we have that  $\mathcal{B}$  is reflexive and strictly convex. By Proposition 1, the reflexivity of  $\mathcal{B}$  ensures the existence of a solution of (3). The uniqueness of the solution can also be obtained by the strict convexity of  $\mathcal{B}$ . That is, the MNI problem (3) has a unique solution  $\hat{f}$ .

Since  $\mathcal{B}$  is smooth, Theorem 8 ensures that there exist coefficients  $\tilde{c}_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that

$$\mathcal{G}(\hat{f}) = \sum_{j \in \mathbb{N}_m} \tilde{c}_j g_j^\sharp. \quad (65)$$

It suffices to identify the Gâteaux derivative  $\mathcal{G}(\hat{f})$  with the dual element  $\hat{f}^\sharp$  of  $\hat{f}$ . The relation between the semi-inner-product and the Gâteaux derivative of the norm  $\|\cdot\|_{\mathcal{B}}$  was given in Giles (1967), that is,

$$\lim_{t \rightarrow 0} \frac{\|g + tf\|_{\mathcal{B}} - \|g\|_{\mathcal{B}}}{t} = \frac{[f, g]_{\mathcal{B}}}{\|g\|_{\mathcal{B}}}, \quad \text{for all } f, g \in \mathcal{B} \text{ and } g \neq 0.$$

This together with (9) and the definition of the dual element leads to  $\mathcal{G}(g) = \frac{g^\sharp}{\|g\|_{\mathcal{B}}}$ , for all  $g \in \mathcal{B} \setminus \{0\}$ . Notice that for  $g = 0$  there holds  $g^\sharp = 0$  and  $\mathcal{G}(g) = 0$ . Hence, we can get a generalized formula as

$$g^\sharp = \|g\|_{\mathcal{B}} \mathcal{G}(g), \quad \text{for all } g \in \mathcal{B}. \quad (66)$$

Substituting this representation with  $g := \hat{f}$  into the left hand side of equation (65), we get that  $\hat{f}^\sharp = \sum_{j \in \mathbb{N}_m} \|\hat{f}\|_{\mathcal{B}} \tilde{c}_j g_j^\sharp$ . Choosing  $c_j := \|\hat{f}\|_{\mathcal{B}} \tilde{c}_j$ ,  $j \in \mathbb{N}_m$ , we obtain the desired representation of  $\hat{f}$  as in statement (1).

The fact that  $\mathcal{B}$  is uniformly convex guarantees that  $\mathcal{B}^*$  is uniformly Fréchet smooth. Thus, the hypotheses of Remark 11 are satisfied. By Remark 11 there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\hat{f} = \rho \mathcal{G}^* \left( \sum_{j \in \mathbb{N}_m} c_j g_j^\sharp \right)$ , with  $\rho := \left\| \sum_{j \in \mathbb{N}_m} c_j g_j^\sharp \right\|_{\mathcal{B}^*}$ . Combining this equation with (66), we get the desired representation of  $\hat{f}$  as in statement (2).  $\blacksquare$

To prove Lemma 15, we need the notion of the dual element of  $\nu \in \mathcal{B}^*$ . Since the dual space  $\mathcal{B}^*$  is also uniformly Fréchet smooth and uniformly convex, there also exists a unique semi-inner-product  $[\cdot, \cdot]_{\mathcal{B}^*} : \mathcal{B}^* \times \mathcal{B}^* \rightarrow \mathbb{R}$  that induces the norm of  $\mathcal{B}^*$ . Furthermore, it was pointed out in Giles (1967) that

$$[f^\sharp, g^\sharp]_{\mathcal{B}^*} = [g, f]_{\mathcal{B}}, \quad \text{for all } f, g \in \mathcal{B}, \quad (67)$$

defines the semi-inner-product on  $\mathcal{B}^*$ . Again, the semi-inner-product  $[\cdot, \cdot]_{\mathcal{B}^*}$  is *not* linear with respect to the second variable. Note that  $\mathcal{B}$  is reflexive. According to the semi-inner-product (67) on  $\mathcal{B}^*$ , we can also define the dual element  $\nu^\sharp \in \mathcal{B}$  of  $\nu \in \mathcal{B}^*$  as

$$\langle \mu, \nu^\sharp \rangle_{\mathcal{B}} := [\mu, \nu]_{\mathcal{B}^*}, \quad \text{for all } \mu \in \mathcal{B}^*. \quad (68)$$

**Proof** [of Lemma 15] On one hand, since  $f^{\sharp\sharp}$  is the dual element of  $f^\sharp \in \mathcal{B}^*$ , for any  $g^\sharp \in \mathcal{B}^*$  there holds  $\langle g^\sharp, f^{\sharp\sharp} \rangle_{\mathcal{B}} = \langle f^{\sharp\sharp}, g^\sharp \rangle_{\mathcal{B}^*} = [g^\sharp, f^\sharp]_{\mathcal{B}^*}$ . On the other hand, for any  $g^\sharp \in \mathcal{B}^*$  there holds  $\langle g^\sharp, f \rangle_{\mathcal{B}} = [f, g]_{\mathcal{B}}$ . Hence, according to (67), we get that  $\langle g^\sharp, f^{\sharp\sharp} \rangle_{\mathcal{B}} = \langle g^\sharp, f \rangle_{\mathcal{B}}$ , for all  $g^\sharp \in \mathcal{B}^*$ , which further implies  $f^{\sharp\sharp} = f$ .  $\blacksquare$

**Proof** [of Corollary 17] The reflexivity and strict convexity of the RKBS  $\mathcal{B}$  ensure that problem (3) has a unique solution  $\hat{f}$ . Since  $\mathcal{B}$  is smooth, Theorem 8 with  $\nu_j := \delta_{x_j}$ ,  $j \in \mathbb{N}_m$ , ensures that there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\mathcal{G}(\hat{f}) = \sum_{j \in \mathbb{N}_m} c_j \delta_{x_j}$ . It follows from the reproducing property that the right-sided reproducing kernel  $K$  provides a closed-form function representation for the point-evaluation functionals. Hence, the above representation coincides with the desired one.  $\blacksquare$

**Proof** [of Theorem 18] As pointed out in Corollary 17, problem (3) has a unique solution  $\hat{f}$ . Since  $\mathcal{B}$  is reflexive, it follows from the strict convexity of  $\mathcal{B}$  that  $\mathcal{B}^*$  is smooth. The hypotheses of Remark 11 are satisfied. Hence, by Remark 11, there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\hat{f} = \rho \mathcal{G}^* \left( \sum_{j \in \mathbb{N}_m} c_j \delta_{x_j} \right)$ , with  $\rho := \left\| \sum_{j \in \mathbb{N}_m} c_j \delta_{x_j} \right\|_{\mathcal{B}^*}$ . By using the reproduction property, we obtain from the above equation the desired formula.  $\blacksquare$

**Proof** [of Theorem 20] We first consider the case when  $\mathbf{y} = 0$ . The MNI problem (3) with  $\mathbf{y} = 0$  has a unique solution  $\hat{\mathbf{x}} = 0$ . On one hand, if  $\hat{\mathbf{x}} = 0$ , there hold  $\hat{\mathbf{x}} \in \mathcal{M}_0$  and the desired representation of  $\hat{\mathbf{x}}$  with  $c_j = 0$ ,  $j \in \mathbb{N}_m$ . On the other hand, suppose that  $\hat{\mathbf{x}} \in \mathcal{M}_0$  and the desired representation holds for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ . Set  $\mathbf{u} := (u_j : j \in \mathbb{N})$ . It follows that  $\langle \mathbf{u}, \hat{\mathbf{x}} \rangle_{\ell_1} = 0$  and

$$\hat{\mathbf{x}} = \|\mathbf{u}\|_\infty \sum_{k \in \mathbb{N}_n} t_k \text{sign}(u_{j_k}) \mathbf{e}_{j_k}, \quad (69)$$

for some  $n \in \mathbb{N}$ ,  $j_k \in \mathbb{N}(\mathbf{u})$ ,  $t_k \in \mathbb{R}_+$ ,  $k \in \mathbb{N}_n$ , with  $\sum_{k \in \mathbb{N}_n} t_k = 1$ . If  $\mathbf{u} = 0$ , equation (69) leads to  $\hat{\mathbf{x}} = 0$ . If  $\mathbf{u} \neq 0$ , by equation (69) we have that

$$\langle \mathbf{u}, \hat{\mathbf{x}} \rangle_{\ell_1} = \|\mathbf{u}\|_\infty \sum_{k \in \mathbb{N}_n} t_k \text{sign}(u_{j_k}) \langle \mathbf{u}, \mathbf{e}_{j_k} \rangle_{\ell_1} = \|\mathbf{u}\|_\infty \sum_{k \in \mathbb{N}_n} t_k |u_{j_k}|. \quad (70)$$

By definition of  $\mathbb{N}(\mathbf{u})$ , there holds  $|u_{j_k}| = \|\mathbf{u}\|_\infty$ , for all  $k \in \mathbb{N}_n$ . Substituting the equations above and the fact that  $\sum_{k \in \mathbb{N}_n} t_k = 1$  into the right hand side of equation (70), we obtain that  $\langle \mathbf{u}, \hat{\mathbf{x}} \rangle_{\ell_1} = \|\mathbf{u}\|_\infty^2$ , which together with  $\langle \mathbf{u}, \hat{\mathbf{x}} \rangle_{\ell_1} = 0$  yields  $\mathbf{u} = 0$ . This is a contradiction.

We prove this result for  $\mathbf{y} \neq 0$  by employing Theorem 12. Note that the MNI problem (3) with  $\mathbf{y} \neq 0$  has no trival solution. Since  $\ell_1(\mathbb{N})$  has  $c_0$  as its pre-dual space, Theorem 12 ensures that  $\hat{\mathbf{x}} \in \ell_1(\mathbb{N})$  is a solution of the MNI problem (3) if and only if  $\hat{\mathbf{x}} \in \mathcal{M}_{\mathbf{y}}$  and there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , such that  $\hat{\mathbf{x}} \in \gamma \partial \|\cdot\|_\infty \left( \sum_{j \in \mathbb{N}_m} c_j \mathbf{u}_j \right)$ , with  $\gamma := \left\| \sum_{j \in \mathbb{N}_m} c_j \mathbf{u}_j \right\|_\infty$ . Substituting the subdifferential formula of Lemma 19 into the right hand side of the above equation and letting  $\mathbf{u} := \sum_{j \in \mathbb{N}_m} c_j \mathbf{u}_j$ , we get the desired representation of  $\hat{\mathbf{x}}$ .  $\blacksquare$

## Appendix D. Proofs for Section 4

**Proof** [of Theorem 21] Suppose that  $\hat{f}$  in the form (16) for some  $\mathbf{c} \in \mathbb{R}^m$  is a solution of (3) with  $\mathbf{y}$ . Substituting (16) into the interpolation condition  $\mathcal{L}(\hat{f}) = \mathbf{y}$ , we have that the vector  $\mathbf{c}$  satisfies (17).

Conversely, we suppose that the vector  $\mathbf{c}$  satisfies the system of equations (17). We first comment that  $\hat{f}$  in the form (16) is in  $\mathcal{B}$  since the operator  $\mathcal{G}_*$  maps  $\mathcal{B}_*$  to  $\mathcal{B}$ . We will prove by Remark 13 that  $\hat{f}$  is a solution of the MNI problem (3) with  $\mathbf{y}$ . Substituting (16) into (17) leads to the interpolation condition  $\mathcal{L}(\hat{f}) = \mathbf{y}$ . Then by Remark 13 and the representation of the adjoint operator  $\mathcal{L}^*$ , we conclude that  $\hat{f} \in \mathcal{B}$  is a solution of the MNI problem (3) with data  $\mathbf{y}$ .  $\blacksquare$

**Proof** [of Corollary 22] By Proposition 21 we have that  $\hat{f}$  in the form (16) is the solution of the MNI problem (3) with  $\mathbf{y}$  if and only if  $\mathbf{c} \in \mathbb{R}^m$  satisfies (17). According to the relation between the semi-inner-product and the Gâteaux derivative of the norm  $\|\cdot\|_{\mathcal{B}}$

$$\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}^*} \mathcal{G}^*(\mathcal{L}^*(\mathbf{c})) = (\mathcal{L}^*(\mathbf{c}))^\sharp, \quad (71)$$

we represent  $\hat{f}$  as in  $\hat{f} = (\mathcal{L}^*(\mathbf{c}))^\sharp$ . Substituting (71) into (17), with noting that  $\nu_k := g_k^\sharp$  for all  $k \in \mathbb{N}_m$ , we have that  $\langle g_k^\sharp, (\mathcal{L}^*(\mathbf{c}))^\sharp \rangle_{\mathcal{B}} = y_k$ ,  $k \in \mathbb{N}_m$ . This together with (68) leads to

(19). That is,  $\hat{f}$  having the form  $\hat{f} = (\mathcal{L}^*(\mathbf{c}))^\sharp$  is a solution of the MNI problem (3) with  $\mathbf{y}$  if and only if  $\mathbf{c} \in \mathbb{R}^m$  is a solution of (19).  $\blacksquare$

To prove Proposition 23, for any  $\mathbf{c} \in \mathbb{R}^m$ , we set  $\phi(\mathbf{c}) := \|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*}$  and  $\psi(\mathbf{c}) := \langle \mathbf{c}, \mathbf{y} \rangle_{\mathbb{R}^m} - 1$ . We also need to describe the chain rule of the subdifferential (Showalter, 1997). Let  $\mathcal{B}_1$  and  $\mathcal{B}_2$  be two real Banach spaces. Supposet that  $\varphi : \mathcal{B}_2 \rightarrow \mathbb{R} \cup \{+\infty\}$  is a convex function and  $\mathcal{T} : \mathcal{B}_1 \rightarrow \mathcal{B}_2$  is a bounded linear operator. If  $\varphi$  is continuous at some point of the range of  $\mathcal{T}$ , then for all  $f \in \mathcal{B}_1$

$$\partial(\varphi \circ \mathcal{T})(f) = \mathcal{T}^* \partial\varphi(\mathcal{T}(f)). \quad (72)$$

**Proof** [of Proposition 23] As a composition of the linear function  $\mathcal{L}^*(\cdot)$  and the convex function  $\|\cdot\|_{\mathcal{B}_*}$ , the function  $\phi$  is convex on  $\mathbb{R}^m$ . Moreover, it is easy to see the convexity and the continuity of the function  $\psi$ . By Lemma 69,  $\hat{\mathbf{c}}$  is a solution of the optimization problem (20) with  $\mathbf{y}$  if and only if  $\langle \hat{\mathbf{c}}, \mathbf{y} \rangle_{\mathbb{R}^m} = 1$  and there exists  $\eta \in \mathbb{R}$  such that  $0 \in \partial\phi(\hat{\mathbf{c}}) + \eta\partial\psi(\hat{\mathbf{c}})$ . Note that  $\phi = \|\cdot\|_{\mathcal{B}_*} \circ \mathcal{L}^*$ ,  $\mathcal{L}^* : \mathbb{R}^m \rightarrow \mathcal{B}_*$  is a bounded linear operator and the norm  $\|\cdot\|_{\mathcal{B}_*}$  is continuous on  $\mathcal{B}_*$ . Then by the chain rule (72) of subdifferentials, we have that  $\partial\phi(\hat{\mathbf{c}}) = \mathcal{L}\partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*\hat{\mathbf{c}})$ . Since  $\psi$  is linear, there holds  $\partial(\psi)(\hat{\mathbf{c}}) = \mathbf{y}$ . Accordingly, we get that  $\partial\phi(\hat{\mathbf{c}}) + \eta\partial\psi(\hat{\mathbf{c}}) = \mathcal{L}\partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*\hat{\mathbf{c}}) + \eta\mathbf{y}$ . It follows that  $\hat{\mathbf{c}}$  is a solution of (20) if and only if  $\langle \hat{\mathbf{c}}, \mathbf{y} \rangle_{\mathbb{R}^m} = 1$  and there exist  $\eta \in \mathbb{R}$  and  $f \in \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*\hat{\mathbf{c}})$  such that  $\mathcal{L}(f) + \eta\mathbf{y} = 0$ . By setting  $\lambda = -\eta$ , we get the desired conclusion.  $\blacksquare$

**Proof** [of Proposition 24] We first suppose that  $\hat{\mathbf{c}} \in \mathbb{R}^m$  is a solution of the minimization problem (20) with  $\mathbf{y}$ . Proposition 23 ensures that  $\langle \hat{\mathbf{c}}, \mathbf{y} \rangle_{\mathbb{R}^m} = 1$  and there exists  $\lambda \in \mathbb{R}$  and  $f \in \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*\hat{\mathbf{c}})$  such that  $\mathcal{L}(f) = \lambda\mathbf{y}$ . It follows from  $f \in \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*\hat{\mathbf{c}})$  that  $\langle \mathcal{L}^*(\hat{\mathbf{c}}), f \rangle_{\mathcal{B}} = \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}$ , which further yields that  $\langle \hat{\mathbf{c}}, \mathcal{L}(f) \rangle_{\mathbb{R}^m} = \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}$ . Substituting  $\mathcal{L}(f) = \lambda\mathbf{y}$  into the latter equation, we have that  $\lambda\langle \hat{\mathbf{c}}, \mathbf{y} \rangle_{\mathbb{R}^m} = \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}$ . This together with  $\langle \hat{\mathbf{c}}, \mathbf{y} \rangle_{\mathbb{R}^m} = 1$  leads to  $\lambda = \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}$ . Set  $\hat{f} := \frac{1}{\lambda}f$ . We will show that  $\hat{f}$  belongs to the intersection in the left side hand of equation (21). Combining inclusion  $f \in \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*\hat{\mathbf{c}})$  with the definition of  $\hat{f}$ , we get that

$$\hat{f} \in \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}^{-1} \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*\hat{\mathbf{c}}). \quad (73)$$

Moreover, equation  $\mathcal{L}(f) = \lambda\mathbf{y}$  leads directly to the interpolation condition  $\mathcal{L}(\hat{f}) = \mathbf{y}$ . That is,  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$ . Consequently, we conclude that  $\hat{f}$  belongs to the intersection in the left hand side of equation (21), which leads to the validity of (21).

Conversely, we suppose that (21) holds. That is, there exists  $\hat{f} \in \mathcal{B}$  satisfying  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and inclusion (73). We will prove by employing Proposition 23 that  $\hat{\mathbf{c}}$  is a solution of the minimization problem (20). By inclusion (73) we get that  $\langle \mathcal{L}^*(\hat{\mathbf{c}}), \hat{f} \rangle_{\mathcal{B}} = 1$ , which yields that  $\langle \hat{\mathbf{c}}, \mathcal{L}(\hat{f}) \rangle_{\mathbb{R}^m} = 1$ . Substituting the interpolation condition  $\mathcal{L}(\hat{f}) = \mathbf{y}$  into the above equation, we have that  $\langle \hat{\mathbf{c}}, \mathbf{y} \rangle_{\mathbb{R}^m} = 1$ . It suffices to verify that there exists  $\lambda \in \mathbb{R}$  and  $f \in \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*\hat{\mathbf{c}})$  such that  $\mathcal{L}(f) = \lambda\mathbf{y}$ . Set  $f := \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*} \hat{f}$  and  $\lambda := \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}$ . Inclusion (73) leads directly to  $f \in \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*\hat{\mathbf{c}})$  and the interpolation condition  $\mathcal{L}(\hat{f}) = \mathbf{y}$  leads to  $\mathcal{L}(f) = \lambda\mathbf{y}$ . Hence, by using Proposition 23 we conclude that  $\hat{\mathbf{c}}$  is a solution of the minimization problem (20) with  $\mathbf{y}$ .  $\blacksquare$

**Proof** [of Theorem 25] We first suppose that  $\hat{f} \in \mathcal{B}$  is a solution of the MNI problem (3) with  $\mathbf{y}$ . Theorem 12 ensures that

$$\hat{f} \in \|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*(\mathbf{c})) \cap \mathcal{M}_{\mathbf{y}} \quad (74)$$

for some  $\mathbf{c} \in \mathbb{R}^m$ . By setting  $\hat{\mathbf{c}} := \frac{\mathbf{c}}{\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*}^2}$ , we get that

$$\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} = \frac{1}{\|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}} \quad \text{and} \quad \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*(\mathbf{c})) = \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*(\hat{\mathbf{c}})). \quad (75)$$

Substituting equations in (75) into the right hand side of inclusion (74), we get the desired inclusion relation, which further leads to (21). Thus, by employing Proposition 24 we conclude that  $\hat{\mathbf{c}}$  is a solution of the minimization problem (20) with  $\mathbf{y}$ .

Conversely, we suppose that  $\hat{\mathbf{c}}$  is a solution of the minimization problem (20) with  $\mathbf{y}$ . Note by Proposition 24 that (21) holds. We also suppose that  $\hat{f}$  is an element satisfying the inclusion relation in this theorem. We will prove by Theorem 12 that  $\hat{f}$  is a solution of the MNI problem (3). Set  $\mathbf{c} := \frac{\hat{\mathbf{c}}}{\|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}^2}$ . It follows that equations in (75) hold. Substituting these two equations into the inclusion relation leads to (74). Hence, Theorem 12 ensures that  $\hat{f}$  is a solution of (3).  $\blacksquare$

**Proof** [of Theorem 26] As has been shown in the proof of Theorem 10 and Remark 11, for the trivial solution  $\hat{f} = 0$ , the coefficients appearing in the solution representations of these theorems are all zeros. Clearly, we have that  $\|\hat{f}\|_{\mathcal{B}} = \left\| \sum_{j \in \mathbb{N}_m} c_j \nu_j \right\|_{\mathcal{B}_*} = 0$ .

It remains to consider the case of having a nontrivial solution  $\hat{f} \neq 0$ . In this case, the function  $\nu := \sum_{j \in \mathbb{N}_m} c_j \nu_j$  is also nonzero. When  $\hat{f}$  satisfies the inclusion relation (11), we get that  $\frac{\hat{f}}{\|\nu\|_{\mathcal{B}_*}} \in \partial\|\cdot\|_{\mathcal{B}_*}(\nu)$ . Equation (8) ensures that  $\frac{\|\hat{f}\|_{\mathcal{B}}}{\|\nu\|_{\mathcal{B}_*}} = 1$ , that is,  $\|\hat{f}\|_{\mathcal{B}} = \|\nu\|_{\mathcal{B}_*}$ . When  $\hat{f}$  satisfies the equality (12), equation  $\|\mathcal{G}^*(\nu)\|_{\mathcal{B}} = 1$ . ensures that  $\|\hat{f}\|_{\mathcal{B}} = \|\nu\|_{\mathcal{B}_*}$ .  $\blacksquare$

## Appendix E. Proofs for Section 5

To prove the theorems in section 5, we first present several useful results. The following relation between the proximity operator of  $\psi : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{+\infty\}$  and its subdifferential can be found in Bauschke and Combettes (2011) and Michelli et al. (2011).

**Lemma 70** *If  $\psi$  is a convex function from  $\mathbb{R}^m$  to  $\mathbb{R} \cup \{+\infty\}$  and  $\mathbf{a} \in \text{dom}(\psi)$ , then*

$$\mathbf{b} \in \partial\psi(\mathbf{a}) \quad \text{if and only if} \quad \mathbf{a} = \text{prox}_{\psi}(\mathbf{a} + \mathbf{b}). \quad (76)$$

In a manner similar to Lemma 70, the proximity operator defined by (25) of a convex function  $\psi$  defined on  $\mathcal{B}$  is intimately related to the subdifferential of  $\psi$ .

**Proposition 71** *Suppose that  $\mathcal{B}$  is a Banach space with the dual space  $\mathcal{B}^*$  and  $\mathcal{H}$  is a Hilbert space. Let  $\mathcal{T}$  be a bounded linear operator from  $\mathcal{B}$  to  $\mathcal{H}$  and  $\mathcal{T}^*$  be its adjoint operator from  $\mathcal{H}$  to  $\mathcal{B}^*$ . If  $\psi : \mathcal{B} \rightarrow \mathbb{R} \cup \{+\infty\}$  is a convex function, then for all  $f \in \text{dom}(\psi)$  and  $g \in \mathcal{B}$*

$$\mathcal{T}^* \mathcal{T}(g) \in \partial\psi(f) \quad \text{if and only if} \quad f = \text{prox}_{\psi, \mathcal{H}, \mathcal{T}}(f + g). \quad (77)$$



**Proof** By definition (25) of the proximity operator on the Banach space  $\mathcal{B}$ , for each  $f \in \text{dom}(\psi)$  and  $g \in \mathcal{B}$ , the equation  $f = \text{prox}_{\psi, \mathcal{H}, \mathcal{T}}(f + g)$  is equivalent to

$$f = \arg \inf \left\{ \frac{1}{2} \|\mathcal{T}(f + g - h)\|_{\mathcal{H}}^2 + \psi(h) : h \in \mathcal{B} \right\}.$$

According to the Fermat rule (Zălinescu, 2002), the above equation holds if and only if  $0 \in \partial \left( \frac{1}{2} \|\mathcal{T}(\cdot - f - g)\|_{\mathcal{H}}^2 + \psi(\cdot) \right) (f)$ . By employing the chain rule (72) of the subdifferential and noting that the subdifferential of the function  $\|\cdot\|_{\mathcal{H}}^2$  at any element in the Hilbert space  $\mathcal{H}$  is a singleton, the inclusion relation above is thus equivalent to  $0 \in \mathcal{T}^* \mathcal{T}(f - f - g) + \partial \psi(f)$ . This inclusion relation is further equivalent to  $\mathcal{T}^* \mathcal{T}(g) \in \partial \psi(f)$ , proving the desired result.  $\blacksquare$

Proposition 71 is a generalization of Lemma 70. We explain this point below. Let  $\mathcal{B}$  be the Euclidean space  $\mathbb{R}^m$  with a norm  $\|\cdot\|$ . For the norm  $\|\cdot\|$ , its dual norm  $\|\cdot\|_{\sharp}$  is defined, for all  $\mathbf{b} \in \mathbb{R}^m$  by  $\|\mathbf{b}\|_{\sharp} := \max\{|\langle \mathbf{a}, \mathbf{b} \rangle_{\mathbb{R}^m}| : \|\mathbf{a}\| = 1, \mathbf{a} \in \mathbb{R}^m\}$ . Accordingly, the dual space  $\mathcal{B}^*$  is identified with  $\mathbb{R}^m$  with the dual norm  $\|\cdot\|_{\sharp}$ . Choose the Hilbert space  $\mathcal{H}$  as  $\mathbb{R}^m$  with the Euclidean norm  $\|\cdot\|_{\mathbb{R}^m}$  and the operator  $\mathcal{T}$  as the identity operator from  $\mathcal{B}$  to  $\mathcal{H}$ . Clearly, the adjoint operator  $\mathcal{T}^*$  is the identity operator from  $\mathcal{H}$  to  $\mathcal{B}^*$ . Hence,  $\mathcal{T}^* \mathcal{T}$  coincides with the identity operator from  $\mathcal{B}$  to  $\mathcal{B}^*$ . In this special case, relation (77) in Proposition 71 reduces to relation (76) in Lemma 70.

The minimization problem (23) involves the composition of the indicator function  $\iota_{\mathbf{y}}$  and the linear operator  $\mathcal{L}$ . We need to compute the subdifferential of the composition function by the chain rule (72) of the subdifferential. However, the pair  $\iota_{\mathbf{y}}$  and  $\mathcal{L}$  does not satisfy the hypothesis of the chain rule (72) of the subdifferential since  $\iota_{\mathbf{y}}$  is not continuous at every point in the range of  $\mathcal{L}$ . Thus, we cannot use the chain rule (72) directly. In the next lemma, we verify that the chain rule for the subdifferential of the composition of these two functions remains valid by using special property of the indicator function  $\iota_{\mathbf{y}}$ .

**Lemma 72** *Suppose that  $\mathcal{B}$  is a Banach space with the dual space  $\mathcal{B}^*$  and  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , are linearly independent. Let  $\mathcal{L}$  be defined by (1) and  $\mathcal{L}^*$  be the adjoint operator. If for a given  $\mathbf{y} \in \mathbb{R}^m$ ,  $\mathcal{M}_{\mathbf{y}}$  is defined by (2) and the indicator function  $\iota_{\mathbf{y}}$  is defined by (22), then for all  $f \in \mathcal{M}_{\mathbf{y}}$*

$$\partial(\iota_{\mathbf{y}} \circ \mathcal{L})(f) = \mathcal{L}^* \partial \iota_{\mathbf{y}}(\mathcal{L}(f)). \quad (78)$$

**Proof** Let  $f \in \mathcal{M}_{\mathbf{y}}$ . By definition (7) of the subdifferential, we have that  $\nu \in \partial(\iota_{\mathbf{y}} \circ \mathcal{L})(f)$  if and only if

$$\iota_{\mathbf{y}}(\mathcal{L}(g)) - \iota_{\mathbf{y}}(\mathcal{L}(f)) \geq \langle \nu, g - f \rangle_{\mathcal{B}}, \quad \text{for all } g \in \mathcal{B}. \quad (79)$$

By the definition of the indicator function  $\iota_{\mathbf{y}}$ , we observe that  $\iota_{\mathbf{y}}(\mathcal{L}(f)) = 0$  and  $\iota_{\mathbf{y}}(\mathcal{L}(g)) = 0$ , for all  $f, g \in \mathcal{M}_{\mathbf{y}}$ . Thus, condition (79) is equivalent to

$$\langle \nu, g - f \rangle_{\mathcal{B}} \leq 0, \quad \text{for all } g \in \mathcal{M}_{\mathbf{y}}. \quad (80)$$

The relation of  $\mathcal{M}_{\mathbf{y}}$  and  $\mathcal{M}_0$  ensures that condition (80) is equivalent to  $\langle \nu, h \rangle_{\mathcal{B}} \leq 0$ , for all  $h \in \mathcal{M}_0$ . Since  $\mathcal{M}_0$  is a subspace of  $\mathcal{B}$ , we can rewrite these inequalities in their equivalent forms  $\langle \nu, h \rangle_{\mathcal{B}} = 0$ , for all  $h \in \mathcal{M}_0$ . That is,  $\nu \in \mathcal{M}_0^{\perp}$ , which guaranteed by Lemma 4 is equivalent to  $\nu \in \text{span} \{\nu_j : j \in \mathbb{N}_m\}$ . Therefore, we conclude that

$$\partial(\iota_{\mathbf{y}} \circ \mathcal{L})(f) = \text{span} \{\nu_j : j \in \mathbb{N}_m\}. \quad (81)$$

On the other hand, clearly, we have that  $\partial\iota_{\mathbf{y}}(\mathbf{y}) = \mathbb{R}^m$ . Hence, for  $\mathcal{L}(f) = \mathbf{y}$  we get that  $\mathcal{L}^*\partial\iota_{\mathbf{y}}(\mathcal{L}(f)) = \mathcal{L}^*\partial\iota_{\mathbf{y}}(\mathbf{y}) = \mathcal{L}^*(\mathbb{R}^m)$ . By the representation of  $\mathcal{L}^*$ , we conclude that

$$\mathcal{L}^*\partial\iota_{\mathbf{y}}(\mathcal{L}(f)) = \text{span}\{\nu_j : j \in \mathbb{N}_m\}. \quad (82)$$

Combining equations (81) and (82), we obtain the desired chain rule (78).  $\blacksquare$

**Proof** [of Theorem 34] By using the Fermat rule together with the chain rule (78) for the subdifferential of the composition function  $\iota_{\mathbf{y}} \circ \mathcal{L}$ , we see that  $\hat{f} \in \mathcal{B}$  is a solution of the minimization problem (23) with  $\mathbf{y}$  if and only if  $0 \in \partial\|\cdot\|_{\mathcal{B}}(\hat{f}) + \mathcal{L}^*\partial\iota_{\mathbf{y}}(\mathcal{L}(\hat{f}))$ . This is equivalent to that there exists  $\mathbf{c} \in \mathbb{R}^m$  such that

$$\mathbf{c} \in \partial\iota_{\mathbf{y}}(\mathcal{L}(\hat{f})) \quad (83)$$

and

$$-\mathcal{L}^*(\mathbf{c}) \in \partial\|\cdot\|_{\mathcal{B}}(\hat{f}). \quad (84)$$

According to the relation (26), we rewrite the inclusion relation (83) as  $\mathcal{L}(\hat{f}) \in \partial\iota_{\mathbf{y}}^*(\mathbf{c})$ . Lemma 70 ensures the equivalence between the inclusion relation above and the first equation in (27). By the assumptions on  $\mathcal{T}$  and noting that  $\mathcal{L}^*(\mathbf{c}) \in \mathcal{V}$ , we rewrite (84) as

$$-(\mathcal{T}^*\mathcal{T})(\mathcal{T}^*\mathcal{T})^{-1}\mathcal{L}^*(\mathbf{c}) \in \partial\|\cdot\|_{\mathcal{B}}(\hat{f}). \quad (85)$$

By employing Proposition 71 with  $\psi := \|\cdot\|_{\mathcal{B}}$ ,  $g := -(\mathcal{T}^*\mathcal{T})^{-1}\mathcal{L}^*(\mathbf{c})$  and  $f := \hat{f}$ , we get that the inclusion relation (85) is equivalent to the second equation in (27). Consequently, we conclude that  $\hat{f} \in \mathcal{B}$  is a solution of the minimization problem (23) with  $\mathbf{y}$  if and only if there exists  $\mathbf{c} \in \mathbb{R}^m$  such that (27).  $\blacksquare$

**Proof** [of Theorem 35] Proposition 24 ensures that  $\hat{\mathbf{c}} \in \mathbb{R}^m$  is a solution of (20) with  $\mathbf{y}$  if and only if there exists  $\hat{f} \in \mathcal{B}$  such that  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and

$$\hat{f} \in \frac{1}{\|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}}\partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*(\hat{\mathbf{c}})). \quad (86)$$

Set  $\mathbf{c} := -\|\hat{f}\|_{\mathcal{B}}\hat{\mathbf{c}}$ . It suffices to verify that  $\hat{f} \in \mathcal{B}$  satisfies  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  and (86) if and only if the pair  $\hat{f}$  and  $\mathbf{c}$  satisfies the fixed-point equations (27). Note that  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  is equivalent to  $\mathcal{L}(\hat{f}) = \mathbf{y}$ . According to the representation of the conjugate function  $\iota_{\mathbf{y}}^*$ , its subdifferential at each  $\mathbf{a} \in \mathbb{R}^m$  is a singleton, that is,  $\partial\iota_{\mathbf{y}}^*(\mathbf{a}) = \{\mathbf{y}\}$ . We then conclude that  $\mathcal{L}(\hat{f}) = \mathbf{y}$  if and only if  $\mathcal{L}(\hat{f}) \in \partial\iota_{\mathbf{y}}^*(\mathbf{c})$ , which guaranteed by Lemma 70 is equivalent to the first equation in (27). Therefore, we have that  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  if and only if  $\hat{f}$  and  $\mathbf{c}$  satisfies the first fixed-point equation in (27).

We next show that  $\hat{f} \in \mathcal{B}$  satisfies (86) if and only if  $\hat{f}$  and  $\mathbf{c}$  satisfies the second fixed-point equation in (27). We rewrite (86) as

$$\|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}\hat{f} \in \partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*(\hat{\mathbf{c}})). \quad (87)$$

Note by the definition of  $\mathbf{c}$  that

$$\partial\|\cdot\|_{\mathcal{B}_*}(\mathcal{L}^*(\hat{\mathbf{c}})) = \partial\|\cdot\|_{\mathcal{B}_*}(-\mathcal{L}^*(\mathbf{c})) \quad (88)$$

and

$$\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} = \|\hat{f}\|_{\mathcal{B}}\|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}. \quad (89)$$

According to (88), we have that (87) holds if and only if

$$\|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*}\hat{f} \in \partial\|\cdot\|_{\mathcal{B}_*}(-\mathcal{L}^*(\mathbf{c})), \quad (90)$$

which guaranteed by (8) is equivalent to

$$\|\hat{f}\|_{\mathcal{B}}\|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}_*} = 1 \quad (91)$$

and

$$\frac{\hat{f}}{\|\hat{f}\|_{\mathcal{B}}} \in \partial\|\cdot\|_{\mathcal{B}_*}(-\mathcal{L}^*(\mathbf{c})). \quad (92)$$

Note by equation (89) that (91) is equivalent to  $\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} = 1$ . Accordingly, we conclude that (90) holds if and only if there hold  $\|-\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} = 1$  and (92). Lemma 9 ensures the equivalence between the latter and inclusion relation (84). As has been shown in the proof of Theorem 34, inclusion relation (84) is equivalent to the second equation in (27). Consequently, we have that  $\hat{f} \in \mathcal{B}$  satisfies (86) if and only if  $\hat{f}$  and  $\mathbf{c}$  satisfies the second fixed-point equation in (27). This completes the proof of this theorem.  $\blacksquare$

To prove Theorems 36 and 37, we also need the following result. Since  $\mathcal{T}_0$  and  $\mathcal{T}_0^*$  are both embedding operator, there holds  $\mathcal{T}_0^*\mathcal{T}_0\mathbf{z} = \mathbf{z}$ , for all  $\mathbf{z} \in \ell_1(\mathbb{N})$ . Hence, as a consequence of Proposition 71, we may get the relation between the proximity operator (29) of  $\psi$  and its subdifferential. Specifically, if  $\psi : \ell_1(\mathbb{N}) \rightarrow \mathbb{R} \cup \{+\infty\}$  is a convex function, then for all  $\mathbf{x} \in \text{dom}(\psi)$  and  $\mathbf{z} \in \ell_1(\mathbb{N})$  that

$$\mathbf{z} \in \partial\psi(\mathbf{x}) \text{ if and only if } \mathbf{x} = \text{prox}_{\psi, \ell_2(\mathbb{N}), \mathcal{T}_0}(\mathbf{x} + \mathbf{z}). \quad (93)$$

As a direct consequence of Lemma 19, we get the following relation between the subdifferential of the norm  $\|\cdot\|_{\infty}$  of  $c_0$  at nonzero  $\mathbf{u} \in c_0$  and that at  $\mathcal{S}(\mathbf{u})$ .

**Lemma 73** *If the truncation operator  $\mathcal{S} : c_0 \rightarrow c_c$  is the truncation operator, then for each nonzero  $\mathbf{u} \in c_0$ ,*

$$\partial\|\cdot\|_{\infty}(\mathbf{u}) = \partial\|\cdot\|_{\infty}(\mathcal{S}(\mathbf{u})). \quad (94)$$

**Proof** It follows from the definition of  $\mathcal{S}$  that for each  $\mathbf{u} \in c_0$ , there holds  $\mathbb{N}(\mathbf{u}) = \mathbb{N}(\mathcal{S}(\mathbf{u}))$ , which together with definition (15) leads to  $\mathcal{V}(\mathbf{u}) = \mathcal{V}(\mathcal{S}(\mathbf{u}))$ . Hence, by Lemma 19 we get that  $\partial\|\cdot\|_{\infty}(\mathbf{u}) = \text{co}(\mathcal{V}(\mathbf{u})) = \text{co}(\mathcal{V}(\mathcal{S}(\mathbf{u}))) = \partial\|\cdot\|_{\infty}(\mathcal{S}(\mathbf{u}))$ , proving the desired result.  $\blacksquare$

By employing the above lemma, we get below a technical lemma, which is useful for establishing the finite dimensional fixed-point equations for a solution of the minimization problem (31).

**Lemma 74** *If  $\mathcal{S} : c_0 \rightarrow c_c$  is the truncation operator, then for all  $\mathbf{x} \in \ell_1(\mathbb{N})$  and all  $\mathbf{u} \in c_0$ ,*

$$\mathbf{u} \in \partial\|\cdot\|_1(\mathbf{x}) \text{ if and only if } \mathcal{S}(\mathbf{u}) \in \partial\|\cdot\|_1(\mathbf{x}). \quad (95)$$

**Proof** By the definition of the subdifferential, we have that  $\partial\|\cdot\|_1(0) = \{\mathbf{v} \in \ell_\infty(\mathbb{N}) : \|\mathbf{v}\|_\infty \leq 1\}$ . This together with the fact that  $\|\mathbf{u}\|_\infty = \|\mathcal{S}(\mathbf{u})\|_\infty$  ensures that (95) holds for  $\mathbf{x} = 0$ . It remains to prove the desired conclusion in the case that  $\mathbf{x} \neq 0$ . By Lemma 9, there holds  $\mathbf{u} \in \partial\|\cdot\|_1(\mathbf{x})$  if and only if  $\frac{\mathbf{x}}{\|\mathbf{x}\|_1} \in \partial\|\cdot\|_\infty(\mathbf{u})$ . Lemma 73 ensures that the last inclusion relation is equivalent to  $\frac{\mathbf{x}}{\|\mathbf{x}\|_1} \in \partial\|\cdot\|_\infty(\mathcal{S}(\mathbf{u}))$ . Again, using Lemma 9 and noting that  $\|\mathcal{S}(\mathbf{u})\|_\infty = \|\mathbf{u}\|_\infty = 1$ , we conclude that the above inclusion relation is equivalent to  $\mathcal{S}(\mathbf{u}) \in \partial\|\cdot\|_1(\mathbf{x})$ , proving the desired result (95).  $\blacksquare$

**Proof** [of Theorem 36] As has been shown in the proof of Theorem 34,  $\hat{\mathbf{x}} \in \ell_1(\mathbb{N})$  is a solution of the minimization problem (31) with  $\mathbf{y}$  if and only if there exists  $\mathbf{c} \in \mathbb{R}^m$  such that

$$\mathbf{c} \in \partial t_{\mathbf{y}}(\mathcal{L}(\hat{\mathbf{x}})) \quad \text{and} \quad -\mathcal{L}^*(\mathbf{c}) \in \partial\|\cdot\|_1(\hat{\mathbf{x}}). \quad (96)$$

By relation (26), the first inclusion relation of (96) has the equivalent form  $\mathcal{L}(\hat{\mathbf{x}}) \in \partial t_{\mathbf{y}}^*(\mathbf{c})$ , which guaranteed by Lemma 70 is equivalent to the first fixed-point equation in (32). Since  $\mathbf{u}_j \in c_0$ ,  $j \in \mathbb{N}_m$ , we have that  $\mathcal{L}^*(\mathbf{c}) \in c_0$ . Hence, by Lemma 74, we conclude that the second inclusion relation of (96) holds if and only if

$$-\mathcal{S}\mathcal{L}^*(\mathbf{c}) \in \partial\|\cdot\|_1(\hat{\mathbf{x}}). \quad (97)$$

Note that  $-\mathcal{S}\mathcal{L}^*(\mathbf{c}) \in \ell_1(\mathbb{N})$ . Relation (93) ensures that the inclusion relation (97) is equivalent to the second fixed-point equation in (32). Consequently, we have that  $\hat{\mathbf{x}} \in \ell_1(\mathbb{N})$  is a solution of the minimization problem (31) with  $\mathbf{y}$  if and only if there exists  $\mathbf{c} \in \mathbb{R}^m$  satisfying (32).  $\blacksquare$

**Proof** [of Theorem 37] By Proposition 24, we have that  $\hat{\mathbf{c}} \in \mathbb{R}^m$  is a solution of the dual problem with  $\mathbf{y}$  if and only if there exists  $\hat{\mathbf{x}} \in \ell_1(\mathbb{N})$  such that  $\hat{\mathbf{x}} \in \|\mathcal{L}^*(\hat{\mathbf{c}})\|_\infty^{-1} \partial\|\cdot\|_\infty(\mathcal{L}^*(\hat{\mathbf{c}})) \cap \mathcal{M}_{\mathbf{y}}$ . It suffices to verify that  $\hat{\mathbf{x}} \in \ell_1(\mathbb{N})$  satisfies the above inclusion relation if and only if the pair  $\hat{\mathbf{x}}$  and  $\mathbf{c} := -\|\hat{\mathbf{x}}\|_1 \hat{\mathbf{c}}$  satisfy (32). As pointed out in the proof of Theorem 35, we conclude that  $\hat{\mathbf{x}} \in \mathcal{M}_{\mathbf{y}}$  if and only if  $\hat{\mathbf{x}}$  and  $\mathbf{c}$  satisfy the first equation in (32). We also have that

$$\hat{\mathbf{x}} \in \|\mathcal{L}^*(\hat{\mathbf{c}})\|_\infty^{-1} \partial\|\cdot\|_\infty(\mathcal{L}^*(\hat{\mathbf{c}})) \quad (98)$$

if and only if  $-\mathcal{L}^*(\mathbf{c}) \in \partial\|\cdot\|_1(\hat{\mathbf{x}})$ . This guaranteed by Lemma 74 is equivalent to  $-\mathcal{S}\mathcal{L}^*(\mathbf{c}) \in \partial\|\cdot\|_1(\hat{\mathbf{x}})$ . By relation (93) and noting that  $-\mathcal{S}\mathcal{L}^*(\mathbf{c}) \in \ell_1(\mathbb{N})$ , the conclusion relation above can be characterized by the second fixed-point equation in (32). Therefore, we get the conclusion that  $\hat{\mathbf{x}}$  satisfies (98) if  $\hat{\mathbf{x}}$  and  $\mathbf{c}$  satisfy the second equation in (32). This completes the proof of this theorem.  $\blacksquare$

To prove Theorem 38, we first present a technical lemma and a proposition.

**Lemma 75** *If  $\mathbf{u} \in c_0$  is nonzero, then for each  $\mathbf{x} \in \partial\|\cdot\|_\infty(\mathbf{u})$ , there hold*

$$\mathbf{x} \in c_c \quad \text{and} \quad \text{supp}(\mathbf{x}) \subseteq \text{supp}(\mathcal{S}(\mathbf{u})). \quad (99)$$

**Proof** Note that for all nonzero element  $\mathbf{u} \in c_0$ , there holds  $\text{supp}(\mathcal{S}(\mathbf{u})) = \mathbb{N}(\mathbf{u})$ . By Lemma 19, each  $\mathbf{x} \in \partial\|\cdot\|_\infty(\mathbf{u})$  is a convex combination of elements of  $\mathcal{V}(\mathbf{u})$  whose supports

are contained in  $\mathbb{N}(\mathbf{u})$ . This leads to the inclusion relation in (99) and thus,  $\mathbf{x} \in c_c$ .  $\blacksquare$

The next proposition ensures that the proximity operator of the norm function  $\|\cdot\|_1$  is a mapping from  $\ell_1(\mathbb{N})$  to  $c_c$ .

**Proposition 76** *If  $\mathcal{T}_0 : \ell_1(\mathbb{N}) \rightarrow \ell_2(\mathbb{N})$  is defined by (28) and the proximity operator  $\text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0}$  is defined by (29) with  $\psi := \|\cdot\|_1$ , then  $\text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0}(\mathbf{x}) \in c_c$ , for all  $\mathbf{x} \in \ell_1(\mathbb{N})$ .*

**Proof** For  $\mathbf{x} := (x_j : j \in \mathbb{N})$ , we let  $\mathbf{y} = (y_j : j \in \mathbb{N}) := \text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0}(\mathbf{x})$ . It follows from equation (30) that  $y_j = \max\{|x_j| - 1, 0\} \text{sign}(x_j)$ , for all  $j \in \mathbb{N}$ . Since  $\mathbf{x} \in \ell_1(\mathbb{N})$ , there exists a positive integer  $N$  such that  $|x_j| < 1$ , for all  $j > N$ . This together with the above equations leads to  $y_j = 0$ , for all  $j > N$ . That is,  $\mathbf{y} \in c_c$ .  $\blacksquare$

**Proof** [of Theorem 38] Proposition 76 ensures that the proximity operator  $\text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0}$  is a mapping from  $\ell_1(\mathbb{N})$  to  $c_c$ . Thus, for any  $\mathbf{s} \in (\mathbb{R}^m, \ell_1(\mathbb{N}))$ , we get that  $\hat{\mathbf{x}} - \mathcal{S}\mathcal{L}^*(\mathbf{c}) \in \ell_1(\mathbb{N})$  and then

$$\text{prox}_{\|\cdot\|_1, \ell_2(\mathbb{N}), \mathcal{T}_0}(\hat{\mathbf{x}} - \mathcal{S}\mathcal{L}^*(\mathbf{c})) \in c_c. \quad (100)$$

On the other hand, the proximity operator  $\text{prox}_{\iota_{\mathbf{y}}^*}$  is a mapping from  $\mathbb{R}^m$  to itself. Note that for any  $\mathbf{s} \in (\mathbb{R}^m, \ell_1(\mathbb{N}))$ , there holds  $\mathbf{c} + \mathcal{L}(\hat{\mathbf{x}}) \in \mathbb{R}^m$ . Therefore, we have that

$$\text{prox}_{\iota_{\mathbf{y}}^*}(\mathbf{c} + \mathcal{L}(\hat{\mathbf{x}})) \in \mathbb{R}^m. \quad (101)$$

Combining (100) with (101), we conclude that  $\mathcal{P} \circ \mathcal{R}$  is an operator from  $(\mathbb{R}^m, \ell_1(\mathbb{N}))$  to  $(\mathbb{R}^m, c_c)$ .

It remains to verify that the fixed-point  $\mathbf{s}$  of operator  $\mathcal{P} \circ \mathcal{R}$  satisfies the assertion of this theorem. Suppose that  $\mathbf{s}$  is a fixed-point of operator  $\mathcal{P} \circ \mathcal{R}$ . That is,  $\hat{\mathbf{x}}$  and  $\mathbf{c}$  satisfy the fixed-point equations (32). According to the proof of Theorem 36, we observe that  $\hat{\mathbf{x}}$  satisfies the second inclusion relation of (96), which guaranteed by Lemma 9 leads to  $\frac{\hat{\mathbf{x}}}{\|\hat{\mathbf{x}}\|_1} \in \partial\|\cdot\|_\infty(-\mathcal{L}^*(\mathbf{c}))$ . By Lemma 75, the above inclusion ensures that  $\mathbf{s}$  satisfies  $\hat{\mathbf{x}} \in c_c$  and  $\text{supp}(\hat{\mathbf{x}}) \subseteq \text{supp}(\mathcal{S}(\mathcal{L}^*(\mathbf{c})))$ .  $\blacksquare$

## Appendix F. Proofs for Section 6

To prove Proposition 40 for the case that  $\mathcal{B}$  has a separable pre-dual space  $\mathcal{B}^*$ , we provide the following lemma which ensures the existence of a bounded minimizing sequence in  $\mathcal{B}$ . For notational convenience, we set

$$\mathcal{R}(f) := \mathcal{Q}_{\mathbf{y}}(\mathcal{L}(f)) + \lambda\varphi(\|f\|_{\mathcal{B}}), \quad \text{for all } f \in \mathcal{B}. \quad (102)$$

**Lemma 77** *Suppose that  $\mathcal{B}$  is a Banach space with the dual space  $\mathcal{B}^*$ ,  $\nu_j \in \mathcal{B}^*$ ,  $j \in \mathbb{N}_m$ , and  $\mathcal{L}$  is defined by (1). Let  $\mathbf{y} \in \mathbb{R}^m$ ,  $\mathcal{Q}_{\mathbf{y}} : \mathbb{R}^m \rightarrow \mathbb{R}_+$ ,  $\varphi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ ,  $\lambda > 0$  be as those appearing in (35) and  $\mathcal{R}$  be defined by (102). If  $\varphi$  is coercive, then there exists a bounded sequence  $f_n, n \in \mathbb{N}$ , in  $\mathcal{B}$  such that*

$$\lim_{n \rightarrow +\infty} \mathcal{R}(f_n) = \inf_{f \in \mathcal{B}} \mathcal{R}(f). \quad (103)$$

**Proof** For any  $\epsilon > 0$ , there exists an element  $g \in \mathcal{B}$  such that  $\inf_{f \in \mathcal{B}} \mathcal{R}(f) \leq \mathcal{R}(g) < \inf_{f \in \mathcal{B}} \mathcal{R}(f) + \epsilon$ . Hence, there exists a sequence  $f_n, n \in \mathbb{N}$ , in  $\mathcal{B}$  satisfying (103). It remains to show that the sequence is bounded. It follows from (103) that  $\{\mathcal{R}(f_n) : n \in \mathbb{N}\}$  is a bounded set. Moreover, by the definition (102) of  $\mathcal{R}$ , we have that  $\mathcal{R}(f_n) \geq \lambda\varphi(\|f_n\|_{\mathcal{B}})$ , for all  $n \in \mathbb{N}$ . This together with the boundedness of the set  $\{\mathcal{R}(f_n) : n \in \mathbb{N}\}$  implies that  $\{\varphi(\|f_n\|_{\mathcal{B}}) : n \in \mathbb{N}\}$  is also a bounded set. By the coercivity of  $\varphi$ , the boundedness of the set  $\{\varphi(\|f_n\|_{\mathcal{B}}) : n \in \mathbb{N}\}$  leads to the boundedness of the sequence  $f_n, n \in \mathbb{N}$ .  $\blacksquare$

With the help of Lemma 77, we prove Proposition 40 as follows.

**Proof** [of Proposition 40] We prove the existence for the case that  $\mathcal{B}$  has a separable pre-dual space  $\mathcal{B}_*$ . Since  $\varphi$  is coercive, by Lemma 77 there exists a bounded sequence  $f_n, n \in \mathbb{N}$ , in  $\mathcal{B}$  satisfying (103) with  $\mathcal{R}$  being defined by (102). It follows from the Banach-Alaoglu theorem that there exists a subsequence  $f_{n_k}, k \in \mathbb{N}$ , weakly\* converges to  $\hat{f} \in \mathcal{B}$ . We shall prove that the weak\* accumulation point  $\hat{f}$  is a solution of the regularization problem (35). This is done by showing that

$$\mathcal{R}(\hat{f}) \leq \liminf_{j \rightarrow +\infty} \mathcal{R}(f_{n_{k_j}}), \quad (104)$$

where  $f_{n_{k_j}}, j \in \mathbb{N}$ , is a subsequence of the sequence  $f_{n_k}, k \in \mathbb{N}$ .

By the definition (102) of  $\mathcal{R}$ , we consider the fidelity term  $\mathcal{Q}_{\mathbf{y}}(\mathcal{L}(\hat{f}))$  and the regularization term  $\varphi(\|\hat{f}\|_{\mathcal{B}})$  separately. We first consider the fidelity term. Since  $\nu_j \in \mathcal{B}_*, j \in \mathbb{N}_m$ , the linear functionals  $\nu_j, j \in \mathbb{N}_m$ , are weakly\* continuous. Hence, we conclude that the linear operator  $\mathcal{L}$  defined by (1) in terms of the linear functionals  $\nu_j, j \in \mathbb{N}_m$ , is weakly\* continuous. The assumption that  $\mathcal{Q}_{\mathbf{y}}$  is lower semi-continuous yields  $\mathcal{Q}_{\mathbf{y}} \circ \mathcal{L}$  is weakly\* lower semi-continuous. Hence, by the weak\* convergence of the sequence  $f_{n_k}, j \in \mathbb{N}$ , we obtain that

$$\mathcal{Q}_{\mathbf{y}}(\mathcal{L}(\hat{f})) \leq \liminf_{j \rightarrow +\infty} \mathcal{Q}_{\mathbf{y}}(\mathcal{L}(f_{n_k})). \quad (105)$$

We now consider the regularization term. Noting that the norm  $\|\cdot\|_{\mathcal{B}}$  is weak\* continuous on  $\mathcal{B}$ , by the weak\* convergence of the sequence  $f_{n_k}, k \in \mathbb{N}$ , we get that

$$\|\hat{f}\|_{\mathcal{B}} \leq \liminf_{k \rightarrow +\infty} \|f_{n_k}\|_{\mathcal{B}}. \quad (106)$$

Let  $f_{n_{k_j}}, j \in \mathbb{N}$ , be the subsequence of the sequence  $f_{n_k}, k \in \mathbb{N}$ , which attains the limit inferior in (106). It follows that  $\|\hat{f}\|_{\mathcal{B}} \leq \lim_{j \rightarrow +\infty} \|f_{n_{k_j}}\|_{\mathcal{B}}$ . Since  $\varphi$  is lower semi-continuity and increasing, we have that

$$\varphi(\|\hat{f}\|_{\mathcal{B}}) \leq \varphi\left(\lim_{j \rightarrow +\infty} \|f_{n_{k_j}}\|_{\mathcal{B}}\right) \leq \liminf_{j \rightarrow +\infty} \varphi(\|f_{n_{k_j}}\|_{\mathcal{B}}). \quad (107)$$

Finally, combining inequalities (105) and (107) yields the inequality (104), which together with (103) leads to  $\mathcal{R}(\hat{f}) \leq \inf_{f \in \mathcal{B}} \mathcal{R}(f)$ . Clearly, this inequality ensures that  $\hat{f}$  is a solution of the regularization problem (35).

In the case that the pre-dual space  $\mathcal{B}_*$  is not separable, the existence of the solution may be proved by the generalized Weierstrass theorem in the weak\* topology.  $\blacksquare$

**Proof** [of Proposition 41] We first prove statement (i). Suppose that  $\hat{g}$  is a solution of the MNI problem (3) for data  $\mathbf{y} := \mathcal{L}(\hat{f})$ . It follows from  $\hat{f} \in \mathcal{M}_{\mathbf{y}}$  with  $\mathbf{y} := \mathcal{L}(\hat{f})$  that

$$\mathcal{L}(\hat{g}) = \mathcal{L}(\hat{f}) \quad \text{and} \quad \|\hat{g}\|_{\mathcal{B}} \leq \|\hat{f}\|_{\mathcal{B}}. \quad (108)$$

On one hand, the equality in (108) further implies that

$$\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\hat{g})) = \mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\hat{f})). \quad (109)$$

On the other hand, since  $\varphi$  is increasing, from the inequality in (108) we have that

$$\varphi(\|\hat{g}\|_{\mathcal{B}}) \leq \varphi(\|\hat{f}\|_{\mathcal{B}}). \quad (110)$$

Combining (109) and (110), with noting that  $\lambda$  is positive, we obtain that  $\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\hat{g})) + \lambda\varphi(\|\hat{g}\|_{\mathcal{B}}) \leq \mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\hat{f})) + \lambda\varphi(\|\hat{f}\|_{\mathcal{B}})$ . This ensures that  $\hat{g}$  is a solution of the regularization problem (35) with given data  $\mathbf{y} := \mathbf{y}_0$ .

We next show statement (ii). Suppose that  $\varphi$  is strictly increasing. Set  $\hat{\mathbf{y}} := \mathcal{L}(\hat{f})$ . It suffices to verify that  $\|\hat{f}\|_{\mathcal{B}} \leq \|f\|_{\mathcal{B}}$ , for all  $f \in \mathcal{M}_{\hat{\mathbf{y}}}$ . On one hand, for all  $f \in \mathcal{M}_{\hat{\mathbf{y}}}$  we have that  $\mathcal{L}(\hat{f}) = \mathcal{L}(f)$ , which leads to

$$\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\hat{f})) = \mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(f)), \quad \text{for all } f \in \mathcal{M}_{\hat{\mathbf{y}}}. \quad (111)$$

On the other hand, since  $\hat{f}$  is a solution of (35) with given data  $\mathbf{y} := \mathbf{y}_0$ , we get that

$$\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\hat{f})) + \lambda\varphi(\|\hat{f}\|_{\mathcal{B}}) \leq \mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(f)) + \lambda\varphi(\|f\|_{\mathcal{B}}), \quad \text{for all } f \in \mathcal{B}. \quad (112)$$

Combining (111) and (112), with noting that  $\lambda$  is positive, we have that  $\varphi(\|\hat{f}\|_{\mathcal{B}}) \leq \varphi(\|f\|_{\mathcal{B}})$  for all  $f \in \mathcal{M}_{\hat{\mathbf{y}}}$ . Since  $\varphi$  is strictly increasing, we get from the above inequalities that  $\|\hat{f}\|_{\mathcal{B}} \leq \|f\|_{\mathcal{B}}$ , for all  $f \in \mathcal{M}_{\hat{\mathbf{y}}}$ . This ensures that  $\hat{f}$  is a solution of the MNI problem (3) for data  $\mathbf{y} := \mathcal{L}(\hat{f})$ . ■

## Appendix G. Proofs for Section 7

**Proof** [of Proposition 42] Suppose that  $\hat{f}$  is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$ . Let  $f_0$  be a solution of the MNI problem (3) with  $\mathbf{y} := \mathcal{L}(\hat{f})$ . Proposition 41 ensures that  $f_0$  is also a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$ . As a solution of the MNI problem,  $f_0$  has the representations as described in Proposition 7. Hence,  $f_0$  satisfies conditions (i)-(iv) of this proposition.

We now consider the case that  $\varphi$  is strictly increasing. Suppose that  $f_0$  is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$ . Statement (2) of Proposition 41 ensures that  $f_0$  is also a solution of the MNI problem (3) with  $\mathbf{y} := \mathcal{L}(f_0)$ . Again by Proposition 7, we get that  $f_0$  satisfies conditions (i)-(iv). Due to the arbitrariness of  $f_0$ , we obtain the desired conclusion. ■

**Proof** [of Theorem 48] Note that the Banach space  $\mathcal{B}$  is smooth. If  $\varphi$  is increasing, by Remark 43 with  $\nu_j := g_j^\sharp$ ,  $j \in \mathbb{N}_m$ , we have that there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  such that  $\mathcal{G}(f_0) = \sum_{j \in \mathbb{N}_m} \hat{c}_j g_j^\sharp$ , for some  $\hat{c}_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ . Substituting (66) with  $g := f_0$  into the above equation and choosing  $c_j := \|f_0\|_{\mathcal{B}} \hat{c}_j$ ,  $j \in \mathbb{N}_m$ , we obtain the desired representation. We now consider the case when  $\varphi$  is strictly increasing. In this case, Remark 43 ensures that every solution  $f_0$  of (35) with  $\mathbf{y} := \mathbf{y}_0$  has the form  $\mathcal{G}(f_0) = \sum_{j \in \mathbb{N}_m} \hat{c}_j g_j^\sharp$ , for some  $\hat{c}_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ . Combining (66) with this equation, we obtain the desired representation for  $c_j := \|f_0\|_{\mathcal{B}} \hat{c}_j$ ,  $j \in \mathbb{N}_m$ . ■

**Proof** [of Theorem 49] Note that the uniformly Fréchet smooth and uniformly convex Banach space  $\mathcal{B}$  has the smooth dual space  $\mathcal{B}^*$ . If  $\varphi$  is increasing, Remark 45 with  $\nu_j := g_j^\sharp$ ,  $j \in \mathbb{N}_m$ , shows that there exists a solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  such that there exist  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , satisfying  $f_0 = \rho \mathcal{G}^* \left( \sum_{j \in \mathbb{N}_m} c_j g_j^\sharp \right)$ , with  $\rho := \left\| \sum_{j \in \mathbb{N}_m} c_j g_j^\sharp \right\|_{\mathcal{B}^*}$ . Substituting (66) with  $\mathcal{B}$  being replaced by  $\mathcal{B}^*$  and  $g$  by  $\sum_{j \in \mathbb{N}_m} c_j g_j^\sharp$  into the above representation leads to the desired representation.

Moreover, if  $\varphi$  is strictly increasing, Remark 45 ensures that every solution  $f_0$  of (35) with  $\mathbf{y} := \mathbf{y}_0$  has the above representation. Arguments similar to those presented above leads to the desired representation. ■

**Proof** [of Theorem 54] Suppose that  $\varphi$  is increasing. If the regularization problem (35) in  $\ell_1(\mathbb{N})$  with  $\mathbf{y} := \mathbf{y}_0$  has  $\hat{\mathbf{x}} = 0$  as a solution, then the trival solution has the desired form for some  $c_j = 0$ ,  $j \in \mathbb{N}_m$ . We consider the case that the regularization problem (35) in  $\ell_1(\mathbb{N})$  with  $\mathbf{y} := \mathbf{y}_0$  has no trival solution. Since  $\ell_1(\mathbb{N})$  has  $c_0$  as its pre-dual space, by Theorem 46 with  $\nu_j := \mathbf{u}_j$ ,  $j \in \mathbb{N}_m$ , there exists a nonzero solution  $\hat{\mathbf{x}}$  of (35) such that  $\hat{\mathbf{x}} \in \|\mathbf{u}\|_\infty \partial \|\cdot\|_\infty(\mathbf{u})$ , for some  $c_j \in \mathbb{R}$ ,  $j \in \mathbb{N}_m$ , where  $\mathbf{u} := \sum_{j \in \mathbb{N}_m} c_j \mathbf{u}_j$ . Substituting the subdifferential formula of Lemma 19 into above equation we get the desired representation.

If  $\varphi$  is strictly increasing, the trival solution  $\hat{\mathbf{x}} = 0$ , provided its existence, has the desired form for some  $c_j = 0$ ,  $j \in \mathbb{N}_m$ . Moreover, Theorem 46 ensures that any nontrival solution satisfies  $\hat{\mathbf{x}} \in \|\mathbf{u}\|_\infty \partial \|\cdot\|_\infty(\mathbf{u})$ . This together with the subdifferential formula of Lemma 19 completes the proof of this theorem. ■

**Proof** [of Theorem 55] Since  $\varphi$  is strictly increasing, Remark 47 ensures that every solution  $f_0$  of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  has the form (39), for some  $\hat{\mathbf{c}} \in \mathbb{R}^m$ . It suffices to show that  $f_0$  in the form (39) is a solution of (35) if and only if  $\hat{\mathbf{c}}$  is a solution of the minimization problem (40). To this end, we define a subset  $\mathcal{A}$  of  $\mathcal{B}$  by  $\mathcal{A} := \{f \in \mathcal{B} : f = \|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} \mathcal{G}_*(\mathcal{L}^*(\mathbf{c})), \mathbf{c} \in \mathbb{R}^m\}$ . Clearly, the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  is equivalent to

$$\inf \{ \mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(f)) + \lambda \varphi(\|f\|_{\mathcal{B}}) : f \in \mathcal{A} \}. \quad (113)$$

Note that each  $f \in \mathcal{A}$  has the form

$$f := \|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*} \mathcal{G}_*(\mathcal{L}^*(\mathbf{c})). \quad (114)$$



Since there holds  $\|\mathcal{G}_*(\mathcal{L}^*(\mathbf{c}))\|_{\mathcal{B}} = 1$ , we have that

$$\|f\|_{\mathcal{B}} = \|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}_*}. \quad (115)$$

Substituting the representation (114) of  $f \in \mathcal{A}$  and the norm representation (115) into the first term and second term of the objective function of the minimization problem (113), respectively, we observe that the minimization problem (113) is equivalent to (40), proving the desired result.  $\blacksquare$

**Proof** [of Corollary 56] We conclude from Theorem 55 that  $f_0$  in the form (39) is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  if and only if  $\hat{\mathbf{c}}$  is a solution of the minimization problem (40). It surfaces to represent  $f_0$  in the form  $f_0 := \sum_{j \in \mathbb{N}_m} \hat{c}_j g_j$  and to reformulate the minimization problem (40) in the form (41). Note that  $\mathcal{H}_* = \mathcal{H}$ . It follows that

$$\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{H}} \mathcal{G}(\mathcal{L}^*(\mathbf{c})) = \mathcal{L}^*(\mathbf{c}), \quad \text{for all } \mathbf{c} \in \mathbb{R}^m. \quad (116)$$

Substituting equation (116) with  $\mathbf{c} := \hat{\mathbf{c}}$  and the representation of  $\mathcal{L}^*$  into (39), we get the desired form of  $f_0$ . Again by equation (116), we rewrite (40) as

$$\inf \{ \mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}\mathcal{L}^*(\mathbf{c})) + \lambda \|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{H}}^2 : \mathbf{c} \in \mathbb{R}^m \}.$$

According to the definition of the Gram matrix  $\mathbf{G}$ , we have that  $\mathcal{L}\mathcal{L}^*(\mathbf{c}) = \mathbf{G}\mathbf{c}$  and  $\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{H}}^2 = \mathbf{c}^\top \mathbf{G}\mathbf{c}$ , for all  $\mathbf{c} \in \mathbb{R}^m$ . Substituting these equations into the above minimization problem leads to (41).  $\blacksquare$

**Proof** [of Remark 57] Note that the assumption that  $\mathcal{Q}_{\mathbf{y}_0}$  is convex ensures the uniqueness of the solution of the minimization problem (41). Because of the linear independence of  $g_j \in \mathcal{H}$ ,  $j \in \mathbb{N}_m$ , the Gram matrix  $\mathbf{G}$  is symmetric and positive definite. Then by the Fermat rule and the chain rule (72),  $\hat{\mathbf{c}}$  is the solution of (41) if and only if  $0 \in \mathbf{G}\partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}\hat{\mathbf{c}}) + 2\lambda\mathbf{G}\hat{\mathbf{c}}$ , which is equivalent to  $-2\lambda\hat{\mathbf{c}} \in \partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}\hat{\mathbf{c}})$ . If  $\mathcal{Q}_{\mathbf{y}_0}$  is non-differentiable, we can characterize the solution of (41) via a fixed-point equation. According to (26), the above inclusion relation holds if and only if  $\mathbf{G}\hat{\mathbf{c}} \in \partial\mathcal{Q}_{\mathbf{y}_0}^*(-2\lambda\hat{\mathbf{c}})$ . Hence, by Lemma 70 we obtain the equivalence between this inclusion relation and the desired fixed-point equation.  $\blacksquare$

**Proof** [of Remark 58] Note that  $\hat{\mathbf{c}}$  is the solution of (41) if and only if it satisfies  $-2\lambda\hat{\mathbf{c}} \in \partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}\hat{\mathbf{c}})$ . If  $\mathcal{Q}_{\mathbf{y}_0}$  is differentiable, then we have that  $\partial\mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}\hat{\mathbf{c}}) = \{\nabla\mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}\hat{\mathbf{c}})\}$ . Substituting this equation into the above inclusion, we obtain the system  $-2\lambda\hat{\mathbf{c}} = \nabla\mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}\hat{\mathbf{c}})$ . If  $\mathcal{Q}_{\mathbf{y}_0}$  has the form (36), then there holds  $\nabla\mathcal{Q}_{\mathbf{y}_0}(\mathbf{G}\hat{\mathbf{c}}) = 2(\mathbf{G}\hat{\mathbf{c}} - \mathbf{y}_0)$ , which together with the above system leads to the linear system  $(\mathbf{G} + \lambda\mathbf{I})\hat{\mathbf{c}} = \mathbf{y}_0$ .  $\blacksquare$

**Proof** [of Corollary 59] Theorem 55 ensures that  $f_0$  with the form (39) is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  if and only if  $\hat{\mathbf{c}}$  is a solution of the minimization problem (40). By making use of the semi-inner-product, we will represent  $f_0$  in the form  $f_0 := \left( \sum_{j \in \mathbb{N}_m} \hat{c}_j g_j^\sharp \right)^\sharp$  and the minimization problem (40) in the form (43). Note that for the uniformly Fréchet smooth and uniformly convex Banach space  $\mathcal{B}$ , its dual space  $\mathcal{B}^*$  is

identified with the pre-dual space  $\mathcal{B}_*$ . Substituting (66) with  $\mathcal{B}$  being replaced by  $\mathcal{B}^*$  and  $g$  by  $\sum_{j \in \mathbb{N}_m} \hat{c}_j g_j^\sharp$  into (39),  $f_0$  may be rewritten as  $f_0 := \left( \sum_{j \in \mathbb{N}_m} \hat{c}_j g_j^\sharp \right)^\sharp$ . Again by (66), we rewrite (40) as

$$\inf \{ \mathcal{Q}_{y_0}(\mathcal{L}(\mathcal{L}^*(\mathbf{c}))^\sharp) + \lambda \varphi(\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}^*}) : \mathbf{c} \in \mathbb{R}^m \}. \quad (117)$$

It follows from (68) that  $\left\langle g_j^\sharp, (\mathcal{L}^*(\mathbf{c}))^\sharp \right\rangle_{\mathcal{B}} = \left[ g_j^\sharp, \mathcal{L}^*(\mathbf{c}) \right]_{\mathcal{B}^*}$ , for all  $j \in \mathbb{N}_m$ , which together with the representations of  $\mathcal{L}^*$  and  $\mathbf{G}_{\text{s.i.p}}$  leads to

$$\mathcal{L}(\mathcal{L}^*(\mathbf{c}))^\sharp = \mathbf{G}_{\text{s.i.p}}(\mathbf{c}). \quad (118)$$

There holds for all  $\mathbf{c} \in \mathbb{R}^m$  that  $\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}^*}^2 = [\mathcal{L}^*(\mathbf{c}), \mathcal{L}^*(\mathbf{c})]_{\mathcal{B}^*} = \sum_{j \in \mathbb{N}_m} c_j \left[ g_j^\sharp, \mathcal{L}^*(\mathbf{c}) \right]_{\mathcal{B}^*}$ . By the definition of the nonlinear operator  $\mathbf{G}_{\text{s.i.p}}$ , the above equation leads to

$$\|\mathcal{L}^*(\mathbf{c})\|_{\mathcal{B}^*}^2 = \mathbf{c}^\top \mathbf{G}_{\text{s.i.p}}(\mathbf{c}). \quad (119)$$

Substituting equations (118) and (119) into the minimization problem (117), we get the equivalent form (43).  $\blacksquare$

**Proof** [of Remark 60] The assumptions about  $\mathcal{Q}_{y_0}$  and  $\varphi$  ensure the uniqueness of the solution of the minimization problem (43). Note that  $\hat{\mathbf{c}} \neq 0$  is the solution of the minimization problem (43) if and only if

$$0 \in \mathcal{L}^* \partial \mathcal{Q}_{y_0} \left( \mathcal{L}(\mathcal{L}^*(\hat{\mathbf{c}}))^\sharp \right) + \lambda \partial(\varphi \circ \|\cdot\|_{\mathcal{B}}) \left( (\mathcal{L}^*(\hat{\mathbf{c}}))^\sharp \right). \quad (120)$$

Since  $\mathcal{Q}_{y_0}$  is differentiable, there holds

$$\partial \mathcal{Q}_{y_0} \left( \mathcal{L}(\mathcal{L}^*(\hat{\mathbf{c}}))^\sharp \right) = \left\{ \nabla \mathcal{Q}_{y_0} \left( \mathcal{L}(\mathcal{L}^*(\hat{\mathbf{c}}))^\sharp \right) \right\}. \quad (121)$$

The linear independence of  $g_j^\sharp$ ,  $j \in \mathbb{N}_m$ , leads to  $\mathcal{L}^*(\hat{\mathbf{c}}) \neq 0$ . Then by the differentiability of  $\varphi$  and equation (66) with  $g := (\mathcal{L}^*(\hat{\mathbf{c}}))^\sharp$ , we have that

$$\partial(\varphi \circ \|\cdot\|_{\mathcal{B}}) \left( (\mathcal{L}^*(\hat{\mathbf{c}}))^\sharp \right) = \left\{ \varphi'(\|(\mathcal{L}^*(\hat{\mathbf{c}}))^\sharp\|_{\mathcal{B}}) / \|(\mathcal{L}^*(\hat{\mathbf{c}}))^\sharp\|_{\mathcal{B}} \mathcal{L}^*(\hat{\mathbf{c}}) \right\}. \quad (122)$$

Note that the two sets in the right hand side of (121) and (122) are singleton. Substituting (121) and (122) into the right hand side of inclusion (120), with noticing that  $\|(\mathcal{L}^*(\hat{\mathbf{c}}))^\sharp\|_{\mathcal{B}} = \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}^*}$ , we get that  $\hat{\mathbf{c}} \neq 0$  is the solution of the minimization problem (43) if and only if  $\hat{\mathbf{c}}$  is the solution of the nonlinear system

$$\mathcal{L}^* \nabla \mathcal{Q}_{y_0}(\mathcal{L}(\mathcal{L}^*(\hat{\mathbf{c}}))^\sharp) + \lambda (\varphi'(\|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}^*}) / \|\mathcal{L}^*(\hat{\mathbf{c}})\|_{\mathcal{B}^*}) \mathcal{L}^*(\hat{\mathbf{c}}) = 0.$$

Combining (118) with (119) and using the linearity of  $\mathcal{L}^*$ , we rewrite the above system as

$$\mathcal{L}^* \left[ \nabla \mathcal{Q}_{y_0}(\mathbf{G}_{\text{s.i.p}}(\hat{\mathbf{c}})) + \lambda [\varphi'((\hat{\mathbf{c}}^\top \mathbf{G}_{\text{s.i.p}}(\hat{\mathbf{c}}))^{1/2}) / (\hat{\mathbf{c}}^\top \mathbf{G}_{\text{s.i.p}}(\hat{\mathbf{c}}))^{1/2}] \hat{\mathbf{c}} \right] = 0.$$

By the linear independence of  $g_j^\sharp$ ,  $j \in \mathbb{N}_m$ , the above system is equivalent to (44).  $\blacksquare$

**Proof** [of Theorem 61] By employing the Fermat rule, we have that  $f_0 \in \mathcal{B}$  is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  if and only if  $0 \in \partial(\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\cdot)) + \lambda\varphi \circ \|\cdot\|_{\mathcal{B}})(f_0)$ . According to the chain rule (72) of the subdifferential, this inclusion relation can be rewritten as  $0 \in \mathcal{L}^*\partial\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(f_0)) + \lambda\partial(\varphi \circ \|\cdot\|_{\mathcal{B}})(f_0)$ . This is equivalent to that there exists  $\hat{\mathbf{c}} \in \mathbb{R}^m$  such that

$$\hat{\mathbf{c}} \in \partial\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(f_0)) \quad \text{and} \quad -\mathcal{L}^*(\hat{\mathbf{c}})/\lambda \in \partial(\varphi \circ \|\cdot\|_{\mathcal{B}})(f_0). \quad (123)$$

Relation (26) ensures that the first inclusion relation of (123) holds if and only if  $\mathcal{L}(f_0) \in \partial\mathcal{Q}_{\mathbf{y}_0}^*(\hat{\mathbf{c}})$ , which is equivalent to the second equation in (45). Since  $-\frac{1}{\lambda}\mathcal{L}^*(\hat{\mathbf{c}}) \in \mathcal{V}$ , we represent the second inclusion relation of (123) as  $(\mathcal{T}^*\mathcal{T})(\mathcal{T}^*\mathcal{T})^{-1}(-\frac{1}{\lambda}\mathcal{L}^*(\hat{\mathbf{c}})) \in \partial(\varphi \circ \|\cdot\|_{\mathcal{B}})(f_0)$ . By Proposition 71, we conclude that the above relation is equivalent to the second equation in (45). Therefore,  $f_0 \in \mathcal{B}$  is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  if and only if there exists  $\hat{\mathbf{c}} \in \mathbb{R}^m$  satisfying the fixed-point equations (45).  $\blacksquare$

**Proof** [of Corollary 62] Theorem 61 ensures that  $f_0 \in \mathcal{B}$  is a solution of the regularization problem (35) with  $\mathbf{y} := \mathbf{y}_0$  if and only if there exists  $\hat{\mathbf{c}} \in \mathbb{R}^m$  satisfying (45). Note that the first equation in (45) is equivalent to the inclusion relation (123). Since  $\mathcal{Q}_{\mathbf{y}_0}$  is differentiable, the subdifferential of  $\mathcal{Q}_{\mathbf{y}_0}$  at  $\mathcal{L}(f_0)$  is the singleton  $\nabla\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(f_0))$ . That is,  $\hat{\mathbf{c}} = \nabla\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(f_0))$ . Substituting this equation into the second equation in (45) leads to (46).  $\blacksquare$

**Proof** [of Theorem 64] As in the proof of Theorem 61,  $\mathbf{x}_0 \in \ell_1(\mathbb{N})$  is a solution of the regularization problem (47) with  $\mathbf{y} := \mathbf{y}_0$  if and only if there exists  $\hat{\mathbf{c}} \in \mathbb{R}^m$  such that

$$\hat{\mathbf{c}} \in \partial\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\mathbf{x}_0)) \quad \text{and} \quad -\mathcal{L}^*(\hat{\mathbf{c}})/\lambda \in \partial\|\cdot\|_1(\mathbf{x}_0). \quad (124)$$

By relation (26) between the subdifferentials of  $\mathcal{Q}_{\mathbf{y}_0}$  and its conjugate  $\mathcal{Q}_{\mathbf{y}_0}^*$ , we rewrite the first inclusion relation of (124) as  $\mathcal{L}(\mathbf{x}_0) \in \partial\mathcal{Q}_{\mathbf{y}_0}^*(\hat{\mathbf{c}})$ , which is equivalent to the first equation in (48). Lemma 74 ensures that the second inclusion relation of (124) is equivalent to  $-\frac{1}{\lambda}\mathcal{S}(\mathcal{L}^*(\hat{\mathbf{c}})) \in \partial\|\cdot\|_1(\mathbf{x}_0)$ . Relation (93) with  $\psi := \|\cdot\|_1$  leads to the equivalence between the above relation and the second equation in (48).  $\blacksquare$

**Proof** [of Corollary 65] By Theorem 64,  $\mathbf{x}_0 \in \ell_1(\mathbb{N})$  is a solution of the regularization problem (47) with  $\mathbf{y} := \mathbf{y}_0$  if and only if there exists  $\hat{\mathbf{c}} \in \mathbb{R}^m$  satisfying (48). As has been shown in the proof of Theorem 64, the first equation in (48) is equivalent to the first inclusion relation of (124). Since  $\mathcal{Q}_{\mathbf{y}_0}$  is differentiable, we have that  $\partial\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\mathbf{x}_0)) = \{\nabla\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\mathbf{x}_0))\}$ . Substituting the above equation into the first inclusion of (124) leads to  $\hat{\mathbf{c}} = \nabla\mathcal{Q}_{\mathbf{y}_0}(\mathcal{L}(\mathbf{x}_0))$ , which together with the second equation in (48) leads to (49).  $\blacksquare$

## References

B. Adcock, V. Antun, and A. C. Hansen. Uniform recovery in infinite-dimensional compressed sensing and applications to structured binary sampling. *Applied and Computational Harmonic Analysis*, 55: 1–40, 2021.

- 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.
- N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68: 337–404, 1950.
- H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2011.
- Y. Benyamini and J. Lindenstrauss. A predual of  $\ell_1$  which is not isomorphic to a  $C(K)$  space. *Israel Journal of Mathematics*, 13: 246–254, 1972.
- B. Bohn, C. Rieger, and M. Griebel. A representer theorem for deep kernel learning. *Journal of Machine Learning Research*, 20: 1–32, 2019.
- C. de Boor and R. E. Lynch. On splines and their minimum properties. *Journal of Mathematics and Mechanics*, 15: 953–969, 1966.
- D. Braess. *Nonlinear Approximation Theory*. Springer Series in Computational Mathematics, vol. 7. Springer-Verlag, Berlin, 1986.
- E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52: 489–509, 2006.
- J. F. Cai, B. Dong, S. Osher, and Z. Shen. Image restoration: total variation, wavelet frames, and beyond. *Journal of the American Mathematical Society*, 25: 1033–1089, 2012.
- P. Chen, J. Huang, and X. Zhang. A primal-dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Problems*, 29: 025011, 2013.
- Y. Chen, J. Huang, S. Li, Y. Lu, and Y. Xu. A content-adaptive unstructured grid based integral equation method with the TV regularization for SPECT reconstruction. *Inverse Problems and Imaging*, 14: 27–52, 2020.
- Z. Chen, C. A. Micchelli, and Y. Xu. *Multiscale methods for Fredholm integral equations*. Cambridge Monographs on Applied and Computational Mathematics, vol. 28. Cambridge University Press, Cambridge, 2015.
- Z. Chen, Y. Xu, and H. Yang. Fast collocation methods for solving ill-posed integral equations of the first kind. *Inverse Problems*, 24: 065007, 2008.
- R. Cheng and Y. Xu. Minimum norm interpolation in the  $\ell_1(\mathbb{N})$  space. *Analysis and Applications*, 19: 21–42, 2021.
- C. K. Chui, F. Deutsch, and J. D. Ward. Constrained best approximation in Hilbert space. *Constructive Approximation*, 6: 35–64, 1990.
- I. Cioranescu. *Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems*. Mathematics and its Applications, vol. 62. Kluwer Academic Publishers Group, Dordrecht, 1990.

- D. D. Cox and F. O’Sullivan. Asymptotic analysis of penalized likelihood and related estimators. *The Annals of Statistics*, 18: 1676–1695, 1990.
- F. Cucker and S. Smale. On the mathematical foundations of learning. *Bulletin of the American Mathematical Society*, 39: 1–49, 2002.
- I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57: 1413–1457, 2004.
- F. Deutsch. *Best Approximation in Inner Product Spaces*. Springer-Verlag, New York, 2001.
- F. Deutsch, V. A. Ubhaya, and Y. Xu. Dual cones, constrained  $n$ -convex  $L_p$ -approximation, and perfect splines. *Journal of Approximation Theory*, 80: 180–203, 1995.
- F. Deutsch, V. A. Ubhaya, J. D. Ward, and Y. Xu. Constrained best approximation in Hilbert space III. Applications to  $n$ -convex functions. *Constructive Approximation*, 12: 361–384, 1996.
- D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52: 1289–1306, 2006.
- I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. Classics in Applied Mathematics, vol. 28. SIAM, Philadelphia, 1999.
- T. Evgeniou, M. Pontil, and T. Poggio. Regularization networks and support vector machines. *Advances in Computational Mathematics*, 13: 1–50, 2000.
- Q. Fan, Y. Jiao, and X. Lu. A primal dual active set algorithm with continuation for compressed sensing. *IEEE Transactions on Signal Processing*, 62: 6276–6285, 2014.
- 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.
- C. Fefferman. A sharp form of Whitney’s extension theorem. *Annals of Mathematics*, 161: 509–577, 2005.
- C. Fefferman. Fitting a  $C^m$ -smooth function to data III. *Annals of Mathematics*, 170: 427–441, 2009.
- C. Fefferman and B. Klartag. Fitting a  $C^m$ -smooth function to data I. *Annals of Mathematics*, 169: 315–346, 2009.
- C. Fefferman and B. Klartag. Fitting a  $C^m$ -smooth function to data II. *Rev. Mat. Iberoamericana*, 25: 49–273, 2009.
- R. J. P. De Figueiredo and G. Chen. PDL $_g$  splines defined by partial differential operators with initial and boundary value conditions. *SIAM Journal on Numerical Analysis*, 27: 519–528, 1990.

- J. R. Giles. Classes of semi-inner-product spaces. *Transactions of the American Mathematical Society*, 129: 436–446, 1967.
- L. Huang, C. Liu, L. Tan, and Q. Ye. Generalized representer theorems in Banach spaces. *Analysis and Applications*, 19: 125–146, 2021.
- Y. Jiang, S. Li, and Y. Xu. A higher-order polynomial method for SPECT reconstruction. *IEEE Transactions on Medical Imaging*, 38: 1271–1283, 2019.
- Q. Jin and X. Lu. A fast nonstationary iterative method with convex penalty for inverse problems in Hilbert spaces. *Inverse Problems*, 30: 045012, 2014.
- P. Jorgensen and F. Tian. Discrete reproducing kernel Hilbert spaces: sampling and distribution of Dirac-masses. *Journal of Machine Learning Research*, 16: 3079–3114, 2015.
- 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.
- J. Krebs, A. K. Louis, and H. Wendland. Sobolev error estimates and a priori parameter selection for semi-discrete Tikhonov regularization. *Journal of Inverse and Ill-Posed Problems*, 17: 845–869, 2009.
- A. Krol, S. Li, L. Shen, and Y. Xu. Preconditioned alternating projection algorithms for maximum a posteriori ECT reconstruction. *Inverse problems*, 28: 115005, 2012.
- A. J. Kurdila and M. Zabaranin. *Convex Functional Analysis*. Systems & Control: Foundations & Applications. Birkhäuser Verlag, Basel, 2005.
- Q. Li, C. A. Micchelli, L. Shen, and Y. Xu. A proximity algorithm accelerated by Gauss-Seidel iterations for L1/TV denoising models. *Inverse Problems*, 28: 095003, 2012.
- Q. Li, L. Shen, Y. Xu, and N. Zhang. Multi-step fixed-point proximity algorithms for solving a class of optimization problems arising from image processing. *Advances in Computational Mathematics*, 41: 387–422, 2015.
- S. Li, J. Zhang, A. Krol, C. R. Schmidlein, L. Vogelsang, L. Shen, E. Lipson, D. Feiglin, and Y. Xu. Effective noise-suppressed and artifact-reduced reconstruction of SPECT data using a preconditioned alternating projection algorithm. *Medical Physics*, 42: 4872–4887, 2015.
- Z. Li, C. A. Micchelli, and Y. Xu. Fixed-point proximity algorithm for minimal norm interpolation. *Applied and Computational Harmonic Analysis*, 49: 328–342, 2020.
- Z. Li, G. Song, and Y. Xu. A fixed-point proximity approach to solving the support vector regression with the group Lasso regularization. *International Journal of Numerical Analysis and Modeling*, 15: 154–169, 2018.
- Z. Li, G. Song, and Y. Xu. A two-step fixed-point proximity algorithm for a class of non-differentiable optimization models in machine learning. *Journal of Scientific Computing*, 81: 923–940, 2019.

- Z. Li, Y. Xu and Q. Ye. Sparse support vector machines in reproducing kernel Banach spaces. *Contemporary computational mathematics—a celebration of the 80th birthday of Ian Sloan*. Vol. 1, 2, 869–887, Springer, Cham, 2018.
- R. Lin, G. Song and H. Zhang. Multi-task Learning in vector-valued reproducing kernel Banach spaces with the  $\ell^1$  norm. *Journal of Complexity*, 63: 101514, 2021.
- J. Lu, L. Shen, C. Xu, and Y. Xu. Multiplicative noise removal in imaging: an exp-model and its fixed-point proximity algorithm. *Applied and Computational Harmonic Analysis*, 41: 518–539, 2016.
- Y. Lu, L. Shen, and Y. Xu. Integral equation models for image restoration: high accuracy methods and fast algorithms. *Inverse Problems*, 26: 045006, 2010.
- R. E. Megginson. *An introduction to Banach space theory*. Graduate Texts in Mathematics, vol. 183. Springer-Verlag, New York, 1998.
- C. A. Micchelli and M. Pontil. A function representation for learning in Banach spaces. In *Learning theory*, 255–269, Lecture Notes in Computer Science 3120, Springer, Berlin, 2004.
- C. A. Micchelli, L. Shen, and Y. Xu. Proximity algorithms for image models: denoising. *Inverse Problems*, 27: 045009, 2011.
- C. A. Micchelli, P. W. Smith, J. Swetits, and J. D. Ward. Constrained  $L_p$  approximation. *Constructive Approximation*, 1: 93–102, 1985.
- C. A. Micchelli, Y. Xu, and H. Zhang. Universal kernels. *Journal of Machine Learning Research*, 7: 2651–2667, 2006.
- J. J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Paris Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, 255: 2897–2899, 1962.
- A. Pinkus. *On  $L_1$ -Approximation*. Cambridge University Press, London, 1989.
- N. G. Polson, J. G. Scott, and B. T. Willard. Proximal algorithms in statistics and machine learning. *Statistical Science*, 30: 559–581, 2015.
- L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60: 259–268, 1992.
- R. Schaback and H. Wendland. Kernel techniques: from machine learning to meshless methods. *Acta Numerica*, 1–97, 2006.
- 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*, 416–426, Springer-Verlag, London, 2001.
- B. Schölkopf and A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, Cambridge, 2002.

- T. Schuster, A. Rieder, and F. Schöpfer. The approximate inverse in action: IV. Semi-discrete equations in a Banach space setting. *Inverse Problems*, 28: 104001, 2012.
- R. E. Showalter. *Monotone Operators in Banach Spaces and Nonlinear Partial Differential Equations*. American Mathematical Society, Providence, 1997.
- S. Smale and D.-X. Zhou. Shannon sampling and function reconstruction from point values. *Bulletin (New Series) of the American Mathematical Society*, 41: 279–305, 2004.
- G. Song and H. Zhang. Reproducing kernel Banach spaces with the  $\ell^1$  norm II: Error analysis for regularized least square regression. *Neural Computation*, 23: 2713–2729, 2011.
- G. Song, H. Zhang, and F. J. Hickernell. Reproducing kernel Banach spaces with the  $\ell_1$  norm. *Applied and Computational Harmonic Analysis*, 34: 96–116, 2013.
- B. K. Sriperumbudur, K. Fukumizu, and G. R. G. Lanckriet. Learning in Hilbert vs. Banach spaces: A measure embedding viewpoint. *Advances in Neural Information Processing Systems (Cambridge)*, MIT Press, 1773–1781, 2011.
- J. J. Swetits, S. E. Weinstein, and Y. Xu. Approximation in  $L_p[0, 1]$  by  $n$ -convex functions. *Numerical Functional Analysis and Optimization*, 11: 167–179, 1990.
- J. J. Swetits, S. E. Weinstein, and Y. Xu. On the characterization and computation of best monotone approximation in  $L_p[0, 1]$  for  $1 \leq p < \infty$ . *Journal of Approximation Theory*, 60: 58–69, 1990.
- J. J. Swetits, S. E. Weinstein, and Y. Xu. Best  $L_p$ -approximation with multiple constraints for  $1 \leq p < \infty$ . *Journal of Approximation Theory*, 65: 90–108, 1991.
- R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B*, 58: 267–288, 1996.
- V. A. Ubhaya and Y. Xu. Constrained  $L_p$ -approximation by generalized  $n$ -convex functions induced by ECT-systems. *Journal of Approximation Theory*, 80: 423–449, 1995.
- M. Unser. Representer theorems for sparsity-promoting  $\ell_1$  regularization. *IEEE Transactions on Information Theory*, 62: 5167–5180, 2016.
- M. Unser. A representer theorem for deep neural networks. *Journal of Machine Learning Research*, 20: 1–30, 2019.
- M. Unser. A unifying representer theorem for inverse problems and machine learning. *Foundations of Computational Mathematics*, 21: 941–960, 2021.
- R. Wang and Y. Xu. Functional reproducing kernel Hilbert spaces for non-point-evaluation functional data. *Applied and Computational Harmonic Analysis*, 46: 569–623, 2019.
- R. Wang and Y. Xu. Regularization in a functional reproducing kernel Hilbert space. *Journal of Complexity*, 66: 101567, 2021.
- H. Wendland. *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics, vol. 17. Cambridge University Press, Cambridge, 2005.



- Y. Xu. *Best Approximation with Geometric Constraints*. Ph.D. Dissertation. Old Dominion University, Norfolk, 1989.
- Y. Xu and Q. Ye. Generalized Mercer kernels and reproducing kernel Banach spaces. *Memoirs of the American Mathematical Society*, 258: 1243, 2019.
- Y. Xu and H. Zhang. Refinable kernels. *Journal of Machine Learning Research*, 8: 2083–2120, 2007.
- Y. Xu and H. Zhang. Refinement of reproducing kernels. *Journal of Machine Learning Research*, 10: 107–140, 2009.
- C. Zălinescu. *Convex Analysis in General Vector Spaces*. World Scientific Publishing Co. Pte. Ltd., New Jersey, 2002.
- H. Zhang, Y. Xu, and J. Zhang. Reproducing kernel Banach spaces for machine learning. *Journal of Machine Learning Research*, 10: 2741–2775, 2009.
- H. Zhang, Y. Xu, and Q. Zhang. Refinement of operator-valued reproducing kernels. *Journal of Machine Learning Research*, 13: 91–136, 2012.
- H. Zhang and J. Zhang. Regularized learning in Banach spaces as an optimization problem: representer theorems. *Journal of Global Optimization*, 54: 235–250, 2012.
- T. Zhang. On the dual formulation of regularized linear systems with convex risks. *Machine Learning*, 46: 91–129, 2002.
- P. Zhao and B. Yu. On model selection consistency of Lasso. *Journal of Machine Learning Research*, 7: 2541–2563, 2006.
- W. Zheng, S. Li, A. Krol, C. R. Schmittlein, X. Zeng, and Y. Xu. Sparsity promoting regularization for effective noise suppression in SPECT image reconstruction. *Inverse Problems*, 35: 115011, 2019.
- J. Zhu, S. Rosset, T. Hastie, and R. Tibshirani. 1-norm support vector machines. *Advances in Neural Information Processing Systems*, 16: 49–56, 2004.