Convex optimization in sums of Banach spaces ☆

Michael Unser *, Shayan Aziznejad

Biomedical Imaging Group, École polytechnique fédérale de Lausanne (EPFL), Station 17, CH-1015, Lausanne, Switzerland

A R T I C L E   I N F O

A B S T R A C T

We characterize the solution of a broad class of convex optimization problems that address the reconstruction of a function from a finite number of linear measurements. The underlying hypothesis is that the solution is decomposable as a finite sum of components, where each component belongs to its own prescribed Banach space; moreover, the problem is regularized by penalizing some composite norm of the solution. We establish general conditions for existence and derive the generic parametric representation of the solution components. These representations fall into three categories depending on the underlying regularization norm: (i) a linear expansion in terms of predefined "kernels" when the component space is a reproducing kernel Hilbert space (RKHS), (ii) a non-linear (duality) mapping of a linear combination of measurement functionals when the component Banach space is strictly convex, and, (iii) an adaptive expansion in terms of a small number of atoms within a larger dictionary when the component Banach space is not strictly convex. Our approach generalizes and unifies a number of multi-kernel (RKHS) and sparse-dictionary learning techniques for compressed sensing available in the literature. It also yields the natural extension of the classical spline-fitting techniques in (semi-)RKHS to the abstract Banach-space setting.

1. Introduction

1.1. From RKHS to Banach spaces

Reproducing kernel Hilbert spaces (RKHS) play a central role in the classical formulations of machine learning, statistical estimation, and the resolution of linear inverse problems [1,2]. They go hand in hand with quadratic (or Tikhonov) regularization and Gaussian processes [3,4]. The popularity of RKHS in machine

learning stems from the fact that the minimization of Hilbertian norms results in parametric solutions that are linear combinations of kernels (basis functions) centered on the data points [1,5–7], a remarkable property that is supported by the celebrated representer theorem [8].

However, recent works that revolve around the concept of sparsity have demonstrated the advantages of considering Banach spaces instead of Hilbert spaces. In particular, compressed sensing relies on the minimization of ℓ_1-norms. Under suitable conditions, this enables the exact recovery of a signal from a limited number of linear measurements [9–14]. Researchers have established representer theorems that explain the sparsifying effect of the ℓ_1-norm [15] and of its variants, including its continuous-domain counterpart: the $\mathcal{M}$-norm (a.k.a. the total-variation norm of a measure) [16–19]. Likewise, we proved in [20] that non-uniform splines of a type that is matched to the regularization operator are universal solutions of linear inverse problems with generalized total-variation regularization. The main difference with the RKHS (or Tikhonov) framework is that the underlying basis functions—or kernels—are selected in an adaptive fashion and are not necessarily placed on the data points [21]. More recently, we have shown that the effect of such minimum-norm regularization could be characterized in full generality, as described in Theorem 1 below [22,23]. The latter is an "abstract" representer theorem that applies to any Banach space $\mathcal{X}'$ (e.g., $\ell_\infty(\mathbb{Z}) = (\ell_1(\mathbb{Z}))'$ or $\mathcal{M}(\mathbb{R}^d) = (C_0(\mathbb{R}^d))'$) identifiable as the dual of some primary Banach space $\mathcal{X}$.

1.2. From sums of RKHS to sums of Banach spaces

It is a known fundamental property that a convex combination (resp., a tensor product) of reproducing kernels retains the desirable reproducing-kernel property (a.k.a. positive-definiteness) [24]. This has prompted researchers to extend the single-kernel Hilbertian methods of machine learning to a whole range of composite problems that involve direct products or (internal) direct sums of RKHS. (It turns out that direct sums and direct products are topologically equivalent, which is the reason why direct-product spaces are sometimes referred to as *external direct sums* [25].) Examples of practical developments that involve direct product/sums of RKHS are: kernel methods for vector-valued data [26,27], multi-kernel learning [28,29], multiscale approximation [30], and semi-parametric models of the form $\widetilde{f} = f + p_0$, where $f \in \mathcal{H}$ (RKHS) and the second component $p_0 \in \text{span}\{p_n\}_{n=1}^{N_0}$ is finite-dimensional [1]. Likewise, the native spaces of variational splines have an inherent direct-sum structure because the underlying regularization functional is a Hilbertian semi-norm [31–34].

While the Banach counterparts of these methods are still lacking for the most part, there is recent evidence that the use of over-complete dictionaries—in particular, unions of bases—is highly advantageous for the resolution of compressed-sensing problems with sparsity constraints [35–41]. In the case where the dictionary is a single basis, there is a direct relation between this type of signal recovery and the kind of ℓ_1-regularization problem mentioned in Section 1.1 [14]. By taking inspiration from the large body of work already available for RKHS, the next promising step is therefore to investigate this type of reconstruction problem from the unifying perspective of an optimization in a sum of Banach spaces.

1.3. Mathematical context

The aim of this paper is to characterize the solution of a broad class of unconstrained-optimization problems that address the recovery of some unknown function f from a finite number of (possibly noisy) samples or, more generally, from a set of M linear measurements $z_m = \langle \nu_m, f \rangle, m = 1, \ldots, M$. Beyond the fact that we leave the classical framework of RKHS, the specificity of our investigation is that the optimization is performed over some Banach space $\mathcal{X}'$ that has a direct-sum structure and/or is equipped with a composite-norm.

To set the stage, we recall the primary results of [22] and [23] and introduce our abstract optimization framework in the form of a single unified theorem.

Theorem 1 *(General Banach representer theorem). Let us consider the following setting:*

- *A dual pair $(\mathcal{X}, \mathcal{X}')$ of Banach spaces.*
- *The analysis subspace $\mathcal{N}_{\boldsymbol{\nu}} = \text{span}\{\nu_m\}_{m=1}^M \subset \mathcal{X}$ with the ν_m being linearly independent.*
- *The linear measurement operator $\boldsymbol{\nu} : \mathcal{X}' \to \mathbb{R}^M : f \mapsto (\langle \nu_1, f \rangle, \dots, \langle \nu_M, f \rangle)$.*
- *The proper, lower-semicontinuous, and convex loss functional $E : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}^+ \cup \{+\infty\}$.*
- *Some arbitrary strictly increasing and convex function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$.*

Then, for any fixed $\boldsymbol{y} \in \mathbb{R}^M$, the solution set of the generic optimization problem

$$S = \arg\min_{f \in \mathcal{X}'} \left(E(\boldsymbol{y}, \boldsymbol{\nu}(f)) + \psi(\|f\|_{\mathcal{X}'}) \right) \tag{1}$$

is nonempty, convex, and weak*-compact.

   When E is strictly convex, or if it imposes the equality constraint $\boldsymbol{y} = \boldsymbol{\nu}(f)$, then any solution $f_0 \in S \subset \mathcal{X}'$ is an $(\mathcal{X}', \mathcal{X})$-conjugate of a common $\nu_0 \in \mathcal{N}_{\boldsymbol{\nu}} \subset \mathcal{X}$, so that $S \subseteq J(\nu_0)$ (see Definition 1). Depending on the type of Banach space, this then results in the following description of the solution(s):

- *If $\mathcal{X}'$ is a Hilbert space and ψ is strictly convex, then the solution is unique and admits the linear representation with parameter $\boldsymbol{a} \in \mathbb{R}^M$ given as*

$$f_0 = \sum_{m=1}^M a_m \varphi_m, \tag{2}$$

  with $\varphi_m = \mathrm{J}_{\mathcal{X}}\{\nu_m\} \in \mathcal{X}'$, where $\mathrm{J}_{\mathcal{X}}$ is the Riesz map $\mathcal{X} \to \mathcal{X}'$.
- *If $\mathcal{X}'$ is a strictly convex Banach space and ψ is strictly convex, then the solution is unique and admits the parametric representation*

$$f_0 = \mathrm{J}_{\mathcal{X}} \left\{ \sum_{m=1}^M a_m \nu_m \right\}, \tag{3}$$

  where $\mathrm{J}_{\mathcal{X}}$ is the (nonlinear) duality operator $\mathcal{X} \to \mathcal{X}'$ (see Definition 1).
- *Otherwise, when $\mathcal{X}'$ is not strictly convex, the solution set is the weak*-closure of the convex hull of its extremal points (see Definition 3) which can all be expressed as*

$$f_0 = \sum_{k=1}^{K_0} c_k e_k \tag{4}$$

  for some $K_0 \leq M$, $c_1, \dots, c_{K_0} \in \mathbb{R}$, where $e_1, \dots, e_{K_0} \in \mathcal{X}'$ are some extremal points of the unit ball $B_{\mathcal{X}'} = \{x \in \mathcal{X} : \|x\|_{\mathcal{X}'} \leq 1\}$.

   The definitions and mathematical background for the interpretation of Theorem 1 are provided in Section 2. The first part of Theorem 1 up to (3) is a retranscription of [42, Theorem 1]. The representation (4) for the case where the solution is non-unique is then deducible from Theorem 3.1 of [23]. Since the latter theorem is more general than what is required here, we are providing an alternative proof of the result in Appendix A. It is important to note that the linear expansion in (4) is adaptive, meaning that the actual choice of K_0 and of the basis functions $e_k \in \mathcal{X}'$ is data-dependent. This is the main difference with the two other cases for which Theorem 1 provides an explicit description of the M-dimensional solution manifold.

1.4. Contributions

While the abstract characterization in Theorem 1 is remarkably general, it is practical only for the cases in which the duality operator $J_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}'$ or the extremal points of the unit ball in $\mathcal{X}'$ are known explicitly, for instance when $\mathcal{X}'$ is a RKHS [24,43,8] or when the underlying norm is a variant of the ℓ_1-norm that promotes sparsity [15,23,19]. In this paper, we are extending the applicability of Theorem 1 by starting from basic building blocks (elementary Banach constituents) and by showing how these can be combined via the use of linear transforms and of direct sums to specify more complex regularization norms that can accommodate mixture models.

In Section 3, we present a refinement of Theorem 1 for the cases where $\mathcal{X}'$ admits the decomposition $\mathcal{X}' = \mathcal{X}'_1 \times \cdots \times \mathcal{X}'_N$ (direct product of Banach spaces) or $\mathcal{X}' = \mathcal{X}'_1 \oplus \cdots \oplus \mathcal{X}'_N$ (direct sum of Banach spaces). The main result there is Theorem 2, which explicitly tells us how the underlying direct-product duality mappings and extremal points can be determined from the knowledge of the same entities for the simpler constituent spaces $\mathcal{X}'_n$. In Section 4, we focus our attention on the direct-sum scenario and illustrate the relevance of our framework to the practice of signal processing and data science. In particular, we present an alternative variational formulation for sparse-dictionary learning and a new representer theorem for mixed-norm regularization problems (Theorem 2).

In Section 5, we extend Theorem 1 by replacing the original regularizing norm by a semi-norm that has a finite-dimensional null space $\mathcal{N}_{\boldsymbol{p}} = \mathrm{span}\{p_1, \ldots, p_{N_0}\} \subset \mathcal{X}'$. The main result expressed by Theorem 3 is that this adds a null-space component p_0 to the generic solution(s) of Theorem 1, which is the desired outcome. At the same time, it reduces the intrinsic dimension of the complementary component $s_0 = f_0 - p_0$ from M to $(M - N_0)$. The mathematical analysis amounts to making sure that the solution exists and to then properly split the problem in order to decouple the determination of the two solution components. The significance of our new Theorem 3 is to show that the traditional techniques of spline approximation [31,32,43,33], which involve semi-reproducing-kernel Hilbert spaces [34], are extendable to Banach spaces in general. Likewise, the non-strictly convex scenario in Theorem 3 is consistent with a number of recent results that have appeared in the literature for sparsity-promoting functionals [20,23,19], although the overlap is only partial due to the generality of our present formulation.

2. Mathematical foundations

A Banach space is a complete normed vector space. It is denoted by $(\mathcal{X}, \|\cdot\|_{\mathcal{X}})$ where $\mathcal{X}$ stands for the vector space and $\|\cdot\|_{\mathcal{X}}$ specifies the underlying norm or, simply, by $\mathcal{X}$ (for short). A Banach space $\mathcal{X}$ has a unique topological dual $\mathcal{X}'$ which is itself a Banach space equipped with the dual norm $\|\cdot\|_{\mathcal{X}'}$ (see (7) below). Formally, an element f of the dual space $\mathcal{X}'$ is a continuous linear functional $f : \mathcal{X} \to \mathbb{R}$. Likewise, since $\mathcal{X}$ is embedded in the bidual space $\mathcal{X}'' = (\mathcal{X}')'$, an element $\nu \in \mathcal{X}$, which is therefore also included in $\mathcal{X}''$, can be viewed as a continuous linear functional $\nu : \mathcal{X}' \to \mathbb{R}$. The bilateral character of this association is described by the duality product

$$\mathcal{X} \times \mathcal{X}' \to \mathbb{R} : (\nu, f) \mapsto \langle \nu, f \rangle_{\mathcal{X} \times \mathcal{X}'} = \langle f, \nu \rangle_{\mathcal{X}' \times \mathcal{X}} \in \mathbb{R}, \tag{5}$$

which is a map that is linear and continuous in both arguments. To avoid notational overload, we shall henceforth drop the subscript in the specification of the duality product under the understanding that the first argument is a linear functional that acts on the second argument; for instance, $\nu : f \mapsto \langle \nu, f \rangle$, where $f \in \mathcal{X}'$ usually also has a concrete identification as a vector or a function. Mathematically, the continuity of the duality product or, equivalently, the continuity of ν (or of f) viewed as a linear functional—is expressed by the generic duality bound

$$\left|\langle \nu, f\rangle\right| \le \|\nu\|_{\mathcal{X}}\|f\|_{\mathcal{X}'}, \tag{6}$$

which holds for any $(\nu, f) \in \mathcal{X} \times \mathcal{X}'$—for more details, refer to [25,44]. The upper bound in (6) is consistent with the definition of the dual norm

$$\|f\|_{\mathcal{X}'} \triangleq \sup_{\nu \in \mathcal{X} \setminus \{0\}} \frac{\langle f, \nu\rangle}{\|\nu\|_{\mathcal{X}}}. \tag{7}$$

In fact, the latter identification suggests that the bound in (6) is tight—a property that is embodied in the fundamental notion of duality mapping [45].

Definition 1 *(Duality mapping).* Let $(\mathcal{X}, \mathcal{X}')$ be a dual pair of Banach spaces. Then, the elements $\nu^* \in \mathcal{X}'$ and $\nu \in \mathcal{X}$ form a $(\mathcal{X}', \mathcal{X})$-conjugate pair if they satisfy:

1. Norm preservation: $\|\nu^*\|_{\mathcal{X}'} = \|\nu\|_{\mathcal{X}}$.
2. Sharp duality bound: $\langle \nu, \nu^*\rangle = \|\nu\|_{\mathcal{X}} \|\nu^*\|_{\mathcal{X}'}$.

For any given $\nu \in \mathcal{X}$, the set of admissible conjugates defines the duality mapping

$$J(\nu) = \{\nu^* \in \mathcal{X}' : \|\nu^*\|_{\mathcal{X}'} = \|\nu\|_{\mathcal{X}} \text{ and } \langle \nu, \nu^*\rangle = \|\nu\|_{\mathcal{X}} \|\nu^*\|_{\mathcal{X}'}\}, \tag{8}$$

which is a nonempty subset of $\mathcal{X}'$. Whenever the duality mapping is single-valued (for instance, when $\mathcal{X}'$ is strictly convex), one also defines the duality operator $\mathrm{J}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}'$, which is such that $\nu^* = \mathrm{J}_{\mathcal{X}}\{\nu\}$.

Definition 2. A Banach space $\mathcal{X}$ (or its associated norm $\|\cdot\|_{\mathcal{X}}$) is said to be *strictly convex* if, for all $f_1, f_2 \in \mathcal{X}$ such that $\|f_1\|_{\mathcal{X}} = \|f_2\|_{\mathcal{X}} = 1$ and $f_1 \ne f_2$, one has that $\|\lambda f_1 + (1 - \lambda)f_2\|_{\mathcal{X}} < 1$ for any $\lambda \in (0, 1)$.

The dual mapping is a powerful mathematical tool that facilitates the investigation of optimization problems in Banach spaces. A primary reference on the topic, which includes the characterization of $\mathrm{J}_{\mathcal{X}}$ for the classical L_p spaces, is [46].

Note that the duality operator $\mathrm{J}_{\mathcal{X}}$ is bijective when $\mathcal{X}$ is reflexive and strictly convex [46,47], in which case $\mathrm{J}_{\mathcal{X}}^{-1} = \mathrm{J}_{\mathcal{X}'} : \mathcal{X}' \to \mathcal{X}'' = \mathcal{X}$. It can therefore be viewed as the natural generalization of the celebrated Riesz map [48,49], which describes the linear isometric mapping of a Hilbert space into its dual. The important difference, however, is that the operator $\mathrm{J}_{\mathcal{X}}$ is generally nonlinear. In fact, it is linear if and only if $\mathcal{X}$ is a Hilbert space, in which case it coincides with the Riesz map $\mathcal{X} \to \mathcal{X}'$ [46,22]. This explains the distinction between the first and second scenarios in Theorem 1: The simplification in (2) occurs because we are able to move the operator inside the sum of (3).

The final statement in Theorem 1, which applies to the cases where the solution is non-unique, involves the notion of extremal points.

Definition 3 *(Extremal points).* Let C be a convex set of a Banach space $\mathcal{X}$. The extremal points of C are the points $f \in C$ such that, if there exist $f_1, f_2 \in C$ and $t \in (0, 1)$ such that $f = tf_1 + (1 - t)f_2$, then it necessarily holds that $f = f_1 = f_2$. The set of these extremal points is denoted by $\mathrm{Ext}(C)$.

While the characterization given by (4) is always valid, it is practical only when the Banach space $\mathcal{X}'$ has a unit ball $B_{\mathcal{X}'}$ with comparatively much fewer extremal points than boundary points. The prototypical case is $\ell_1(\mathbb{Z})$, whose extremal points $\mathrm{Ext}(B_{\ell_1(\mathbb{Z})}) = \{\pm\delta[\cdot - m]\}_{m \in \mathbb{Z}}$ (the signed Kronecker impulses shifted by m) are indexable, while its boundary points $\{u[\cdot] \in \ell_1(\mathbb{Z}) : \|u\|_{\ell_1} = 1\}$ are uncountable. The extremal points of $B_{\mathcal{X}'}$ can then be interpreted as the elements of a constrained dictionary. This means that f_0 in (4)

will be constructed by adaptively selecting a few elements e_k (with $K_0 \ll M$ when the solution is strongly regularized) in a dictionary of preferred elementary solutions. This is a very popular approach in compressed sensing [9,10,12,50].

Our final tool is a transformation mechanism that generates application-specific Banach spaces from some primary ones whose basic properties (e.g., the duality mapping and extremal points of the unit ball) are known.

Proposition 1 *(Isometric isomorphism). Let $\mathcal{X}$ be a primary Banach space and $\mathrm{T} : \mathcal{X} \to \mathrm{T}(\mathcal{X})$ a linear operator that is injective on $\mathcal{X}$. Then, we have the following properties:*

1. *The space $\mathcal{Y} = \mathrm{T}(\mathcal{X}) = \{y = \mathrm{T}x : x \in \mathcal{X}\}$, equipped with the norm $\|y\|_{\mathcal{Y}} = \|\mathrm{T}^{-1}\{y\}\|_{\mathcal{X}}$, is a Banach space that is isometrically isomorphic to $\mathcal{X}$. In other words, the operators $\mathrm{T} : \mathcal{X} \to \mathcal{Y}$ and $\mathrm{T}^{-1} : \mathcal{Y} \to \mathcal{X}$ are isometries.*
2. *The continuous dual of $\mathcal{Y} = \mathrm{T}(\mathcal{X})$ is $\mathcal{Y}' = \mathrm{T}^{-1*}(\mathcal{X}')$, equipped with the norm $\|y^*\|_{\mathcal{Y}'} = \|\mathrm{T}^*\{y^*\}\|_{\mathcal{X}'}$.*
3. *The elements $y^* \in \mathcal{Y}'$ and $y \in \mathcal{Y} = \mathrm{T}(\mathcal{X})$ form a conjugate pair if and only if $x^* = \mathrm{T}^*\{y^*\} \in \mathcal{X}'$ and $x = \mathrm{T}^{-1}\{y\} \in \mathcal{X}$ are themselves $(\mathcal{X}', \mathcal{X})$-Banach conjugates.*
4. *The element $u \in \mathcal{Y}$ is an extremal point of the unit ball in $\mathcal{Y} = \mathrm{T}(\mathcal{X})$ if and only if $u = \mathrm{T}\{e\}$, where $e \in \mathcal{X}$ is an extremal point of the unit ball in $\mathcal{X}$.*
5. *If $\mathcal{X}$ is a Hilbert space, then the spaces $\mathcal{Y} = \mathrm{T}(\mathcal{X})$ and $\mathcal{Y}' = \mathrm{T}^{-1*}(\mathcal{X}')$ are Hilbert spaces as well. The corresponding Riesz map is $\mathrm{J}_{\mathcal{Y}} = \mathrm{T}^{-1*}\mathrm{J}_{\mathcal{X}}\mathrm{T}^{-1} : \mathcal{Y} \to \mathcal{Y}'$, where $\mathrm{J}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X}' = \mathrm{J}_{\mathcal{X}}(\mathcal{X})$ is the Riesz map of the primary space.*

Proof. The hypothesis that T is injective on $\mathcal{X}$ implies the existence of a linear map T^{-1} (inverse operator) such that $\mathrm{T}^{-1}\mathrm{T}\{x\} = x$ for all $x \in \mathcal{X}$. Since T is linear and one-to-one, the functional $y \mapsto \|\mathrm{T}^{-1}y\|_{\mathcal{X}}$ is a *bona fide* norm on $\mathcal{Y}$. Moreover, from the definition of the $\mathcal{Y}$-norm, we have that

$$\|\mathrm{T}\{x_m\} - \mathrm{T}\{x_n\}\|_{\mathcal{Y}} = \|\mathrm{T}\{x_m - x_n\}\|_{\mathcal{Y}} = \|\mathrm{T}^{-1}\mathrm{T}\{x_m - x_n\}\|_{\mathcal{X}} = \|x_m - x_n\|_{\mathcal{X}}, \qquad (9)$$

for any $x_m, x_n \in \mathcal{X}$. Together with the bijectivity of T, we deduce that T is an isomorphism between $\mathcal{X}$ and $\mathcal{Y}$. Hence, $\mathcal{Y}$ inherits the topological structure of $\mathcal{X}$. This proves that $\mathcal{Y}$ is indeed a Banach space.

The other properties are immediate consequences of the underlying isometry and the definition of the adjoint, which translate into

$$\langle x_1^*, x_2 \rangle_{\mathcal{X}' \times \mathcal{X}} = \langle x_1^*, \mathrm{T}^{-1}\mathrm{T}\{x_2\} \rangle_{\mathcal{X}' \times \mathcal{X}} = \langle \mathrm{T}^{-1*}\{x_1^*\}, \mathrm{T}\{x_2\} \rangle_{\mathcal{Y}' \times \mathcal{Y}} = \langle y_1^*, y_2 \rangle_{\mathcal{Y}' \times \mathcal{Y}}$$

for any $(x_1^*, x_2) \in \mathcal{X}' \times \mathcal{X}$.

In particular, if $\mathcal{X}$ is a Hilbert space with inner product $(\cdot, \cdot)_{\mathcal{X}}$, then $x^* = \mathrm{J}_{\mathcal{X}}\{x\} \in \mathcal{X}'$ so that $\langle x^*, x \rangle_{\mathcal{X}' \times \mathcal{X}} = \|x\|_{\mathcal{X}}^2 = (x, x)_{\mathcal{X}}$. It follows that $\mathcal{Y} = \mathrm{T}(\mathcal{X})$ is a Hilbert space equipped with the inner product $(y_1, y_2)_{\mathcal{Y}} = (\mathrm{T}^{-1}y_1, \mathrm{T}^{-1}y_2)_{\mathcal{X}}$. Correspondingly, the dual space $\mathcal{Y}' = \mathrm{T}^{-1*}(\mathcal{X}')$ is the Hilbert space equipped with the inner product $(y_1^*, y_2^*)_{\mathcal{Y}'} = (\mathrm{T}^*y_1^*, \mathrm{T}^*y_2^*)_{\mathcal{X}'}$. Moreover, we have that $(y_1, y_2)_{\mathcal{Y}} = \langle \mathrm{J}_{\mathcal{Y}}\{y_1\}, y_2 \rangle_{\mathcal{Y}' \times \mathcal{Y}} = (\mathrm{J}_{\mathcal{Y}}\{y_1\}, \mathrm{J}_{\mathcal{Y}}\{y_2\})_{\mathcal{Y}'}$, the underlying duality operator (Riesz map) being $\mathrm{J}_{\mathcal{Y}} = \mathrm{T}^{-1*}\mathrm{J}_{\mathcal{X}}\mathrm{T}^{-1} : \mathcal{Y} \to \mathcal{Y}'$. $\quad\square$

3. Composite norms and direct-sum spaces

In order to offer flexibility in the specification of direct-product or direct-sum topologies, we introduce the finite-dimensional space $\mathcal{Z} = (\mathbb{R}^N, \|\cdot\|_{\mathcal{Z}})$. The underlying norm is said to be *monotone* if

$$\|(a_1, \ldots, a_N)\|_{\mathcal{Z}} \leq \|(b_1, \ldots, b_N)\|_{\mathcal{Z}}$$

whenever $0 \leq |a_n| \leq |b_n|$ for each $n = 1, \ldots, N$, and, *absolute* if $\|\mathbf{z}\|_{\mathcal{Z}} = \|(z_n)\|_{\mathcal{Z}} = \|(|z_n|)\|_{\mathcal{Z}}$ for any $\mathbf{z} \in \mathbb{R}^N$. It is also known that a norm is monotone if and only if it is absolute [51, Theorem 2]. For instance, the latter property is obviously satisfied for $\|\cdot\|_{\mathcal{Z}} = \|\cdot\|_p$ with $p \geq 1$, as well as for any weighted version thereof. Moreover, the dual of an absolute norm is again absolute [51, Theorem 1]. Given a series $\mathcal{X}_1, \ldots, \mathcal{X}_N$ of Banach spaces, we then write $(\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)_{\mathcal{Z}}$ for the direct-product space equipped with the composite norm

$$\|(x_1, \ldots, x_N)\| = \|(\|x_1\|_{\mathcal{X}_1}, \ldots, \|x_N\|_{\mathcal{X}_N})\|_{\mathcal{Z}}. \tag{10}$$

The construction is straightforward as the direct-product space automatically inherits the Banach property of its components.

Likewise, one can construct (internal) direct-sum spaces via the summation of complemented Banach constituents.

Definition 4. A series $\mathcal{X}_1, \ldots, \mathcal{X}_N$ of Banach subspaces of $\mathcal{X}$ is said to be complemented if $\mathcal{X} = \mathcal{X}_1 + \cdots + \mathcal{X}_N = \{x = x_1 + \cdots + x_N : x_n \in \mathcal{X}_n, n = 1, \ldots, N\}$ (as a set) and $\mathcal{X}_{n_1} \cap \sum_{n \neq n_1} \mathcal{X}_n = \{0\}$ when $n_1 = 1, \ldots, N$.

In that scenario, any $x \in \mathcal{X}$ has a unique representation as $x = x_1 + \cdots + x_N$ with $x_n = \mathrm{Proj}_{\mathcal{X}_n}\{x\} \in \mathcal{X}_n$, where $\mathrm{Proj}_{\mathcal{X}_n} : \mathcal{X} \to \mathcal{X}_n$ is the corresponding projection operator. We then designate $\mathcal{X} = (\mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_N)_{\mathcal{Z}}$ as the (internal) direct-sum space equipped with the norm

$$\|x\|_{\mathcal{X}} = \|(\|\mathrm{Proj}_{\mathcal{X}_1}\{x\}\|_{\mathcal{X}_1}, \ldots, \|\mathrm{Proj}_{\mathcal{X}_N}\{x\}\|_{\mathcal{X}_N})\|_{\mathcal{Z}}. \tag{11}$$

We observe that (11) is compatible with (10) because $\mathrm{Proj}_{\mathcal{X}_{n_1}} : \mathcal{X} \to \mathcal{X}_{n_1}$ is such that

$$\mathrm{Proj}_{\mathcal{X}_{n_1}}\{x_n\} = \begin{cases} x_{n_1}, & \text{for } n = n_1 \\ 0, & \text{otherwise} \end{cases}$$

for any $x_n \in \mathcal{X}_n$. This identification, together with the unicity of the sum decomposition, implies that $(\mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_N)_{\mathcal{Z}}$ is a Banach space that is isometrically isomorphic to $(\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)_{\mathcal{Z}}$.

Lemma 1. *Let $(\mathcal{X}_1', \mathcal{X}_1), \ldots, (\mathcal{X}_N', \mathcal{X}_N)$ be a series of dual pairs of Banach spaces and $\|\cdot\|_{\mathcal{Z}}$ a norm on $\mathbb{R}^N$ that is absolute. Then, we have the following properties:*

1. *The continuous dual of $\mathcal{X} = (\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)_{\mathcal{Z}}$ is the direct-product space $\mathcal{X}' = (\mathcal{X}_1' \times \cdots \times \mathcal{X}_N')_{\mathcal{Z}'}$.*
2. *The elements $y = (y_1, \ldots, y_N) \in \mathcal{X}'$ and $x = (x_1, \ldots, x_N) \in \mathcal{X}$ form a conjugate pair if and only if $y_n = \alpha_n x_n^*$, where $x_n^* \in \mathcal{X}_n'$ is a Banach conjugate of $x_n \in \mathcal{X}_n$ and $\alpha_n \in \mathbb{R}^+$ is given by*

$$\alpha_n = \begin{cases} \frac{z_n^*}{\|x_n\|_{\mathcal{X}_n}} > 0, & x_n \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

   with $\mathbf{z}^* = (z_n^*) \in \mathcal{Z}'$ a Banach conjugate of $\mathbf{z} = (\|x_1\|_{\mathcal{X}_1}, \ldots, \|x_N\|_{\mathcal{X}_N}) \in \mathcal{Z}$.
3. *The element $e = (e_1, \ldots, e_N) \in \mathcal{X}$ is an extremal point of the unit ball in $\mathcal{X}$ if and only if $(\|e_1\|_{\mathcal{X}_1}, \ldots, \|e_N\|_{\mathcal{X}_N})$ is an extremal point of the unit ball in $\mathcal{Z}$, and for each $1 \leq n \leq N$ with $e_n \neq 0$, $\frac{e_n}{\|e_n\|_{\mathcal{X}_n}}$ is an extremal point of the unit ball of $\mathcal{X}_n$.*
4. *If the $\mathcal{X}_n$ are complemented Banach subspaces of the (sum) space $\mathcal{X}_{\mathrm{sum}}$, then the continuous dual of $\mathcal{X}_{\mathrm{sum}} = (\mathcal{X}_1 \oplus \cdots \oplus \mathcal{X}_N)_{\mathcal{Z}}$ is the direct-sum Banach space $\mathcal{X}_{\mathrm{sum}}' = (\mathcal{X}_1' \oplus \cdots \oplus \mathcal{X}_N')_{\mathcal{Z}'}$, which is isometrically isomorphic to the direct-product space $\mathcal{X}'$ in Item 1. Consequently, the properties in Item 2 and 3 also apply, with the convention that $x_n = \mathrm{Proj}_{\mathcal{X}_n}\{x\}$ and $y_n = \mathrm{Proj}_{\mathcal{X}_n'}\{y\}$ for $n = 1, \ldots, N$.*

Proof. An element $y = (y_1, \ldots, y_N)$ of $\mathcal{X}'$ is identified with the linear functional

$$x = (x_1, \ldots, x_N) \mapsto \langle y, x \rangle_{\mathcal{X}' \times \mathcal{X}} = \sum_{n=1}^{N} \langle y_n, x_n \rangle_{\mathcal{X}'_n \times \mathcal{X}_n}. \tag{13}$$

The first property is a basic result in the theory of Banach spaces [25, Theorem 1.10.13] when the outer norm is Euclidean with $\mathcal{Z} = \mathcal{Z}' = (\mathbb{R}^N, \|\cdot\|_2)$. The present setting is more general so that we need to prove that the dual norm of $y = (y_1, \ldots, y_N) \in \mathcal{X}'$ is precisely

$$\|y\|_{\mathcal{X}'} = \sup_{\|x\|_{\mathcal{X}} = 1} \langle y, x \rangle = \left\| (\|y_1\|_{\mathcal{X}'_1}, \ldots, \|y_N\|_{\mathcal{X}'_N}) \right\|_{\mathcal{Z}'}. \tag{14}$$

Since the spaces $(\mathcal{X}'_n, \mathcal{X}_n)$ form dual pairs, we have the generic duality inequalities

$$\langle y_n, x_n \rangle_{\mathcal{X}'_n \times \mathcal{X}_n} \leq |\langle y_n, x_n \rangle_{\mathcal{X}'_n \times \mathcal{X}_n}| \leq \|y_n\|_{\mathcal{X}'_n} \|x_n\|_{\mathcal{X}_n} \tag{15}$$

with equality if and only if $y_n = \alpha_n x_n^*$ for some $\alpha_n \in \mathbb{R}^+$. This implies that, for any $(y, x) \in \mathcal{X}' \times \mathcal{X}$, we have that

$$\langle y, x \rangle_{\mathcal{X}' \times \mathcal{X}} = \sum_{n=1}^{N} \langle y_n, x_n \rangle_{\mathcal{X}'_n \times \mathcal{X}_n} \leq \sum_{n=1}^{N} |\langle y_n, x_n \rangle_{\mathcal{X}'_n \times \mathcal{X}_n}| \leq \sum_{n=1}^{N} \|y_n\|_{\mathcal{X}'_n} \|x_n\|_{\mathcal{X}_n} \tag{16}$$

Likewise, by setting $\boldsymbol{y} = (\|y_1\|_{\mathcal{X}'_1}, \ldots, \|y_N\|_{\mathcal{X}'_N}) \in \mathcal{Z}'$ and $\boldsymbol{z} = (\|x_1\|_{\mathcal{X}_1}, \ldots, \|x_N\|_{\mathcal{X}_N}) \in \mathcal{Z}$, we write the complementary duality inequality

$$\sum_{n=1}^{N} \|y_n\|_{\mathcal{X}'_n} \|x_n\|_{\mathcal{X}_n} = \langle \boldsymbol{y}, \boldsymbol{z} \rangle_{\mathcal{Z}' \times \mathcal{Z}} \leq |\langle \boldsymbol{y}, \boldsymbol{z} \rangle_{\mathcal{Z}' \times \mathcal{Z}}| \leq \|\boldsymbol{y}\|_{\mathcal{Z}'} \|\boldsymbol{z}\|_{\mathcal{Z}}. \tag{17}$$

By observing that $\|\boldsymbol{z}\|_{\mathcal{Z}} = \|x\|_{\mathcal{X}}$ and combining these inequalities, we get that

$$\langle y, x \rangle_{\mathcal{X}' \times \mathcal{X}} \leq \sum_{n=1}^{N} \|y_n\|_{\mathcal{X}'_n} \|x_n\|_{\mathcal{X}_n} \leq \|\boldsymbol{y}\|_{\mathcal{Z}'} \|\boldsymbol{x}\|_{\mathcal{Z}} = \|\boldsymbol{y}\|_{\mathcal{Z}'} \|x\|_{\mathcal{X}}, \tag{18}$$

which shows that $\|y\|_{\mathcal{X}'}$ is upper-bounded by $\|\boldsymbol{y}\|_{\mathcal{Z}'} = \left\| (\|y_1\|_{\mathcal{X}'_1}, \ldots, \|y_N\|_{\mathcal{X}'_N}) \right\|_{\mathcal{Z}'}$. To prove that we actually have $\|y\|_{\mathcal{X}'} = \|\boldsymbol{y}\|_{\mathcal{Z}'}$, for any $\epsilon > 0$, we need to find $x_\epsilon \in \mathcal{X}$ with $\|x_\epsilon\|_{\mathcal{X}} = 1$ such that

$$\langle y, x_\epsilon \rangle_{\mathcal{X}' \times \mathcal{X}} \geq \|\boldsymbol{y}\|_{\mathcal{Z}'} - \epsilon.$$

By definition of the dual norm $\|\cdot\|_{\mathcal{Z}'}$, we have that

$$\|\boldsymbol{y}\|_{\mathcal{Z}'} = \sup_{\substack{\boldsymbol{\alpha} \in \mathbb{R}^N \\ \|\boldsymbol{\alpha}\|_{\mathcal{Z}} \leq 1}} \boldsymbol{y}^T \boldsymbol{\alpha}. \tag{19}$$

Since $\mathbb{R}^N$ is a finite-dimensional vector-space, the unit ball $B_{\mathcal{Z}} = \{\alpha \in \mathbb{R}^N : \|\boldsymbol{\alpha}\|_{\mathcal{Z}} \leq 1\}$ is compact. Hence, there exists a vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N) \in B_{\mathcal{Z}}$ that attains the supremum in (19). In other words,

$$\|\boldsymbol{y}\|_{\mathcal{Z}'} = \boldsymbol{y}^T \boldsymbol{\alpha} = \sum_{n=1}^{N} \|y_n\|_{\mathcal{X}'_n} \alpha_n.$$

Similarly, for any $\epsilon > 0$, the definition of the dual norm implies the existence of unit-norm elements $x_n \in \mathcal{X}_n$ for $n = 1, \ldots, N$ such that

$$\langle y_n, x_n \rangle_{\mathcal{X}'_n \times \mathcal{X}_n} \geq \|y_n\|_{\mathcal{X}'_n} - \frac{2\epsilon}{N(\alpha_n^2 + 1)}. \tag{20}$$

We then set $x_\epsilon = (\alpha_1 x_1, \ldots, \alpha_N x_N) \in \mathcal{X}$ and observe that

$$\|x_\epsilon\|_{\mathcal{X}} = \|(\|\alpha_1 x_1\|_{\mathcal{X}_1}, \ldots, \|\alpha_N x_N\|_{\mathcal{X}_N})\|_{\mathcal{Z}} = \|(|\alpha_1|, \ldots, |\alpha_N|)\|_{\mathcal{Z}} = 1.$$

Based on (20) and the inequality $\frac{\alpha}{\alpha^2+1} \leq \frac{1}{2}$ for all $\alpha \in \mathbb{R}$, we then deduce that

$$\langle y, x_\epsilon \rangle_{\mathcal{X}' \times \mathcal{X}} = \sum_{n=1}^{N} \langle y_n, \alpha_n x_n \rangle_{\mathcal{X}'_n \times \mathcal{X}_n} \geq \sum_{n=1}^{N} \alpha_n \left(\|y_n\|_{\mathcal{X}'_n} - \frac{2\epsilon}{N(\alpha_n^2 + 1)} \right)$$

$$= \sum_{n=1}^{N} \alpha_n \|y_n\|_{\mathcal{X}'_n} - \frac{2\epsilon}{N} \sum_{n=1}^{N} \frac{\alpha_n}{\alpha_n^2 + 1} = \|y\|_{\mathcal{Z}'} - \frac{2\epsilon}{N} \sum_{N=1}^{N} \frac{\alpha_n}{\alpha_n^2 + 1} \geq \|y\|_{\mathcal{Z}'} - \epsilon,$$

which, in light of the inequality $\|y\|_{\mathcal{X}'} \leq \|y\|_{\mathcal{Z}'}$, allows us to conclude that $\|y\|_{\mathcal{X}'} = \|y\|_{\mathcal{Z}'}$.

To prove the second property, we observe that $y \in \mathcal{X}'$ and $x \in \mathcal{X}$ form a conjugate pair if and only if an equality occurs in both (16) and (17). Inequalities (15) and (16) are saturated if and only if $y_n = \alpha_n x_n^*$, $\alpha_n \in \mathbb{R}^+$, and (x_n^*, x_n) form a $(\mathcal{X}'_n, \mathcal{X}_n)$-conjugate pair. The saturation of (17) with $\|y\|_{\mathcal{X}'} = \|y\|_{\mathcal{Z}'} = \|z\|_{\mathcal{Z}} = \|x\|_{\mathcal{X}}$ is then equivalent to $y = z^* = (z_1^*, \ldots, z_N^*)$. Under the assumption that $x_n \neq 0$, this yields $\alpha_n = \frac{z_n^*}{\|x_n^*\|_{\mathcal{X}'_n}}$, which is the announced result since $\|x_n^*\|_{\mathcal{X}'_n} = \|x_n\|_{\mathcal{X}_n}$.

The third property is due to Dowling and Saejung [52] under the assumption that the $\|\cdot\|_{\mathcal{Z}}$-norm is absolute and monotone in the positive orthant; in other words, when the condition $0 \leq a_n \leq b_n$ for $n = 1, \ldots, N$ implies that $\|a\|_Z \leq \|b\|_Z$. By invoking Bauer's theorem [51], we are able to drop the (redundant) assumption of monotonicity since it is implied by the absoluteness property.

The last statement is a direct consequence of the isometric isomorphism between $\mathcal{X}_{\text{sum}} = (\mathcal{X}_1 \oplus \ldots, \oplus \mathcal{X}_N)_{\mathcal{Z}}$ and $\mathcal{X} = (\mathcal{X}_1 \times \ldots, \times \mathcal{X}_N)_{\mathcal{Z}}$. □

In particular, if $\|\cdot\|_{\mathcal{Z}} = \|\cdot\|_2$ is the usual Euclidean norm, then z^* in Property 2 is unique and coincides with z, which implies that the Banach conjugate of $x = (x_1, \ldots, x_N) \in \mathcal{X}$ is simply $x^* = (x_1^*, \ldots, x_N^*) \in \mathcal{X}'$.

The combination of these preparatory results and Theorem 1 allows us to deduce the following.

Theorem 2 *(Representer theorem for direct-product spaces). If the space $\mathcal{X}'$ in Theorem 1 has a direct-product decomposition as $\mathcal{X}' = (\mathcal{X}'_1 \times \cdots \times \mathcal{X}'_N)_{\mathcal{Z}'}$ with predual $\mathcal{X} = (\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)_{\mathcal{Z}}$, where $(\mathcal{X}'_1, \mathcal{X}_1), \ldots, (\mathcal{X}'_N, \mathcal{X}_N)$ are dual pairs of Banach spaces and both E and ψ are strictly convex, then the solutions $f_0 = (f_{0,1}, \ldots, f_{0,N}) \in S \subset \mathcal{X}'$ of the optimization problem (1) are $(\mathcal{X}', \mathcal{X})$-Banach conjugates of a common*

$$\nu_0 = (\nu_{0,1}, \ldots, \nu_{0,N}) = \sum_{m=1}^{M} a_m \nu_m,$$

where $\nu_m = (\nu_{m,1}, \ldots, \nu_{m,N}) \in \mathcal{X}$ with $\nu_{m,n} \in \mathcal{X}_n$ and a suitable set of coefficients $a \in \mathbb{R}^M$.

Moreover, depending of the properties of the underlying Banach constituents, the solution components $f_{0,n} \in \mathcal{X}'_n$ have the following characterization with predefined scaling constants

$$\alpha_n = \begin{cases} \frac{y_n}{y_n^*} > 0, & y_n \neq 0 \\ 0, & \text{otherwise}, \end{cases} \tag{21}$$

where $\boldsymbol{y} = (\|f_{0,1}\|_{\mathcal{X}_1'}, \ldots, \|f_{0,N}\|_{\mathcal{X}_N'})$ and $\boldsymbol{y}^* = \mathrm{J}_{\mathcal{Z}'}\{\boldsymbol{y}\}$:

- If $\mathcal{X}_n'$ is a Hilbert space and $\mathcal{Z}'$ is strictly convex, then $f_{0,n}$ is unique and admits the linear representation

$$f_{0,n} = \alpha_n \sum_{m=1}^{M} a_m \varphi_{m,n} \tag{22}$$

  with $\varphi_{m,n} = \mathrm{J}_{\mathcal{X}_n}\{\nu_{m,n}\} \in \mathcal{X}_n'$, where $\mathrm{J}_{\mathcal{X}_n}$ is the Riesz map $\mathcal{X}_n \to \mathcal{X}_n'$.
- If $\mathcal{X}_n'$ is a strictly convex Banach space and $\mathcal{Z}'$ is strictly convex, then the solution component is unique and admits the parametric representation

$$f_{0,n} = \alpha_n \mathrm{J}_{\mathcal{X}_n}\left\{ \sum_{m=1}^{M} a_m \nu_{m,n} \right\} \tag{23}$$

  where $\mathrm{J}_{\mathcal{X}_n}$ is the (nonlinear) duality operator $\mathcal{X}_n \to \mathcal{X}_n'$.
- If $\mathcal{X}_n'$ is a non-strictly convex Banach space, then the subcomponent solution set $S|_{\mathcal{X}_n'}$ is the weak*-closure of the convex hull of its extremal points, which can all be expressed as

$$f_{0,n} = \sum_{k=1}^{K_0} c_{k,n} e_{k,n}, \tag{24}$$

  where $e_{1,n}, \ldots, e_{K_0,n} \in \mathcal{X}_n'$ are some extremal points of the unit ball in $\mathcal{X}_n'$ and $c_{1,n}, \ldots, c_{K_0,n} \in \mathbb{R}$ some appropriate weights; the (minimal) number of atoms $K_0 \leq M$ is common to all the components associated with non-reflexive Banach spaces.

In the particular case where $\|\cdot\|_{\mathcal{Z}'} = \|\cdot\|_1$, (24) can be replaced by

$$f_{0,n} = \sum_{k=1}^{K_n} c_{k,n} e_{k,n} \tag{25}$$

with $\sum_{n=1}^{N} K_n \leq M$. In addition, (23) (resp. (22)) remains valid for the components for which the space $\mathcal{X}_n'$ is strictly convex (resp., Hilbertian), with the caveat that the solution is no longer guaranteed to be unique; this, then, contributes a degenerate version of (25) with $K_n = 1$, $c_{1,n} = \|f_{0,n}\|_{\mathcal{X}_n'}$, and $e_{1,n} = f_{0,n}/\|f_{0,n}\|_{\mathcal{X}_n'}$.

Proof. The existence of solutions $f_0 \in \mathcal{X}'$ and the property that $S \subseteq J_{\mathcal{X}}(\nu_0)$ for some $\nu_0 = \sum_{m=1}^{M} a_m \nu_m \in \mathcal{N}_{\boldsymbol{\nu}}$ is ensured by Theorem 1. We then proceed in three steps.

(i) Constant value of $\psi(\|f_0\|_{\mathcal{X}'})$ for all $f_0 \in S$.
The key here is the strict convexity of $f \mapsto E(\boldsymbol{y}, \boldsymbol{\nu}(f))$ together with the convexity of $f \mapsto \psi(\|f\|_{\mathcal{X}'})$. By applying a standard argument (by contradiction) that uses the convexity of S, we show that there exist two constants C_1 and C_2 such that $E(\boldsymbol{y}, \boldsymbol{\nu}(f_0)) = C_1$ and $\psi(\|f_0\|_{\mathcal{X}'}) = C_2$ for all $f_0 \in S$ (see, for instance, the last part of the proof in [21, Appendix B]). By invoking the strict convexity of E, this then implies that all solutions share the same measurement vector $\boldsymbol{z}_0 = \boldsymbol{\nu}(f_0)$. Likewise, when ψ is strictly convex, we readily deduce that $\|f_0\|_{\mathcal{X}'}$ takes a constant value.

(ii) Uniqueness of $\|f_{0,n}\|_{\mathcal{X}_n'}$ in the strictly-convex case.
To show that $\|f_{0,n}\|_{\mathcal{X}_n'} = y_n$ holds for all $f_0 \in S$, we suppose that there exists another solution $\widetilde{f}_0 \in S$ such that $\|\widetilde{f}_0\|_{\mathcal{X}'} = \|f_0\|_{\mathcal{X}'}$ and $\|\widetilde{f}_{0,n}\|_{\mathcal{X}_n'} = \widetilde{y}_n$ with $\widetilde{\boldsymbol{y}} \neq \boldsymbol{y}$. Since S is convex, $\lambda \widetilde{f}_0 + (1-\lambda)f_0$ with any $\lambda \in (0,1)$ must also be a solution with associated norm $\|\lambda \widetilde{f}_0 + (1-\lambda)f_0\|_{\mathcal{X}'} \leq \|\lambda \widetilde{\boldsymbol{y}} + (1-\lambda)\boldsymbol{y}\|_{\mathcal{Z}'}$, by the

triangle inequality. However, the norm equality $\|\widetilde{f_0}\|_{\mathcal{X}'} = \|\widetilde{\boldsymbol{y}}\|_{\mathcal{Z}'} = \|\boldsymbol{y}\|_{\mathcal{Z}'}$ and the strict-convexity of $\|\cdot\|_{\mathcal{Z}'}$ (see Definition 2) implies that $\|\lambda\widetilde{\boldsymbol{y}} + (1-\lambda)\boldsymbol{y}\|_{\mathcal{Z}'} < \|\boldsymbol{y}\|_{\mathcal{Z}'} = \|f_0\|_{\mathcal{X}'}$, which results in a contradiction.

(iii) Generic form of the solution component $f_{0,n}$.

We assume that $y_n = \|f_{0,n}\|_{\mathcal{X}'} \neq 0$; otherwise, we simply have that $f_{0,n} = 0$. From Property 2 in Lemma 1, we know that $f_0 = (f_{0,1}, \ldots, f_{0,N})$ and $\nu_0 = (\nu_{0,1}, \ldots, \nu_{0,N})$ form a conjugate pair if and only if there exists $f_{0,n}^* \in \mathrm{J}_{\mathcal{X}_n'}(f_{0,n})$ such that $\nu_{0,n} = (y_n^*/y_n)f_{0,n}^*$, where $\boldsymbol{y}^* = (y_1^*, \ldots, y_N^*) = \mathrm{J}_{\mathcal{X}'}\{\boldsymbol{y}\}$.

When $\mathcal{X}_n'$ is strictly convex, the duality mapping is single-valued. The representations in (22) and (24) then directly follow from the primary expansion $\nu_{0,n} = \sum_{m=1}^{M} a_m \nu_{m,n}$ and the homogeneity property of the duality mapping expressed as $\mathrm{J}_{\mathcal{X}}\{\alpha\nu\} = \alpha\mathrm{J}_{\mathcal{X}}\{\nu\}$ for any $\nu \in \mathcal{X}$ and $\alpha \in \mathbb{R}^+$ (see [46]).

Since S is convex and weak*-compact, we can invoke the Krein-Milman theorem, which states that S is the closure of the convex hull of its extremal points. The same holds true for the convex set $S|_{\mathcal{X}_n'}$ (the restriction of S on $\mathcal{X}_n'$) with $\mathrm{Ext}(S|_{\mathcal{X}_n'}) \subseteq \mathrm{Ext}(S)|_{\mathcal{X}_n'}$. By recalling that all points $f_0 \in \mathrm{Ext}(S)$ can be represented as $f_0 = (f_{0,1}, \ldots, f_{0,N}) = \sum_{k=1}^{K_0} c_k e_k$, where $e_k = (e_{k,1}, \ldots, e_{k,N}) \in \mathrm{Ext}(B_{\mathcal{X}'})$ and $K_0 \leq M$ (by Theorem 1), we obtain that

$$f_{0,n} = \sum_{k=1}^{K_0} c_k \|e_{k,n}\|_{\mathcal{X}_n'} \widetilde{e}_{k,n}, \tag{26}$$

where $\widetilde{e}_{k,n} = e_{k,n}/\|e_{k,n}\|_{\mathcal{X}'}$ are extremal points of the unit ball in $\mathcal{X}_n'$ (by Lemma 1, Property 3). The announced statement with $c_{k,n} = \|e_{k,n}\|_{\mathcal{X}_n'} c_k$ then follows from the property that (26) is valid for all $f_{0,n} \in \mathrm{Ext}(S)|_{\mathcal{X}_n'} \supseteq \mathrm{Ext}(S|_{\mathcal{X}_n'})$. In fact, Property 3 in Lemma 1 tells us that the subset of points $f_{0,n} \in \mathrm{Ext}(S|_{\mathcal{X}_n'})$ are those for which $\boldsymbol{e}_k = \boldsymbol{y}/\|\boldsymbol{y}\|_{\mathcal{Z}'}$ are extremal points of the unit ball in $\mathcal{Z}'$. In particular, when $\|\cdot\|_{\mathcal{Z}'} = \|\cdot\|_1$ (outer ℓ_1-norm), the $\boldsymbol{e}_k$ all have the binary form $(0, 0, \ldots, \pm 1, 0, \ldots)$ with a single active coefficient at $n = n_k$, which then yields (25). $\quad\square$

The outcome of Theorem 2 is that the generic form of the solution in Theorem 1 is essentially transferred to the direct-product components, with the distribution of the relative energy being controlled by the outer norm $\|\cdot\|_{\mathcal{Z}'}$. The effect of the ℓ_1-norm is significant in that respect because it acts as a threshold that selectively blocks certain solution components and lets others through.

We wish to highlight the fact that the arguments for the proof of Theorem 2 (as well as Theorem 1) involves neither a calculus of variations nor a recourse to the sophisticated machinery of Fréchet derivatives and subgradients. It only requires the Hahn-Banach theorem and the characterization of the configurations that saturate the underlying duality inequalities.

4. Convex optimization in sums of Banach spaces

The techniques that we describe next are relevant to inverse problems for which the solution f_0 can be decomposed into a sum of components that have distinct smoothness and/or sparsity properties. The solution then lives in a sum of Banach spaces. Beside the reconstruction of f_0 from the noisy measurement $\boldsymbol{y} = \boldsymbol{\nu}(f) + \boldsymbol{\epsilon}$, we are now faced with the additional challenge of disambiguating the individual components of the solution.

Let $\mathcal{X}_1', \ldots, \mathcal{X}_N'$ be a series of Banach spaces whose elements are indexed over the same domain. We then define the sum space

$$\mathcal{X}_1' + \cdots + \mathcal{X}_N' = \{f = f_1 + f_2 + \cdots + f_N : f_n \in \mathcal{X}_n', n = 1, \ldots, N\}.$$

Given a linear measurement operator $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_M) : \mathcal{X}_1' + \cdots + \mathcal{X}_N' \to \mathbb{R}^M$ with $\nu_n \in \cap_{n=1}^{N}\mathcal{X}_n$ and a set of measurements $\boldsymbol{y} \in \mathbb{R}^M$, we are then interested in the study of the solvability of the convex optimization problem

$$S = \arg \min_{(f_n)_{n=1}^N : f_n \in \mathcal{X}_n'} \left(E\Big(\boldsymbol{y}, \boldsymbol{\nu}(\sum_{n=1}^N f_n)\Big) + \psi\Big(\big\|(\|f_1\|_{\mathcal{X}_1'}, \dots, \|f_N\|_{\mathcal{X}_N'})\big\|_{\mathcal{Z}'}\Big) \right) \tag{27}$$

where the functions E and ψ are the same as in Theorem 1, while $\|\cdot\|_{\mathcal{Z}'}$ is a suitable norm that controls the coupling of the components. The idea here is to segregate the components f_n by favoring some "regularized" solutions $f_0 = (f_{0,n})_{n=1}^N$ such that the $\|f_{0,n}\|_{\mathcal{X}_n'}$ are small in an appropriate sense. Problem (27) is generally well defined. Its solution can be obtained as a special case of Theorem 2. To see this, it suffices to invoke the linearity of ν_m, which yields

$$\nu_m\Big(\sum_{n=1}^N f_n\Big) = \sum_{n=1}^N \langle \nu_m, f_n \rangle_{\mathcal{X}_n \times \mathcal{X}_n'} = \langle \widetilde{\nu}_m, f \rangle_{\mathcal{X} \times \mathcal{X}'}$$

with $f = (f_1, \dots, f_N) \in \mathcal{X}' = (\mathcal{X}_1' \times \cdots \times \mathcal{X}_N')_{\mathcal{Z}'}$, and $\widetilde{\nu}_m = (\nu_m, \dots, \nu_m) \in \mathcal{X} = (\mathcal{X}_1 \times \cdots \times \mathcal{X}_N)_{\mathcal{Z}}$. The multicomponent optimization problem (27) is therefore equivalent to (1) with $\mathcal{X}'$ being a direct-product space and the specific choice of a "replicated" measurement operator $\widetilde{\boldsymbol{\nu}} = (\boldsymbol{\nu}, \dots, \boldsymbol{\nu})$. Consequently, we get the general form of the solution by simple substitution of $\nu_{m,n}$ by ν_m in Theorem 2. We shall now illustrate the power of the approach by considering special cases that are motivated by applications.

4.1. Multicomponent learning in RKHS/RKBS

In the classical supervised learning (or regression) setting [53], one is given a series of data points $(\boldsymbol{x}_m, y_m) \in \mathbb{R}^d \times \mathbb{R}$, $m = 1, \dots, M$. The goal is to determine a function $f : \mathbb{R}^d \to \mathbb{R}$ such that $f(\boldsymbol{x}_m) \approx y_m$ for all m without overfitting the data, which is the reason why one generally imposes some regularization on the solution.

We make the link with our framework by considering the "sampling" functionals $\boldsymbol{\nu} = (\delta(\cdot - \boldsymbol{x}_1), \dots, \delta(\cdot - \boldsymbol{x}_M))$ with $\delta(\cdot - \boldsymbol{x}_m) : f \mapsto f(\boldsymbol{x}_m)$ (Dirac impulse shifted by $\boldsymbol{x}_m$), which is such that $\boldsymbol{\nu}(f) = (f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_M))$. To enable the sequential handling of data, one quantifies the goodness of fit with some additive loss functional of the form $\sum_{m=1}^M E(y_m, f(\boldsymbol{x}_m))$ with $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+ \cup \{+\infty\}$, the simplest case being the least-squares criterion with $E(y_m, f(\boldsymbol{x}_m)) = |y_m - f(\boldsymbol{x}_m)|^2$. The traditional form of regularization is the squared Hilbertian norm $\psi(\|f\|_{\mathcal{X}'}) = \|f\|_{\mathcal{H}}^2$ with $\psi(\cdot) = |\cdot|^2$ (strictly convex) in the reproducing-kernel Hilbert space $\mathcal{X}' = \mathcal{H}$.

Definition 5 *(See [24])*. A Hilbert space $\mathcal{H}$ of functions on $\mathbb{R}^d$ is called a reproducing kernel Hilbert space (RKHS) if its dual $\mathcal{H}'$ is such that $\delta(\cdot - \boldsymbol{x}) \in \mathcal{H}'$ for any $\boldsymbol{x} \in \mathbb{R}^d$. Then, the unique representer $r_{\mathcal{H}}(\cdot, \boldsymbol{x}) = \mathrm{J}_{\mathcal{H}'}\{\delta(\cdot - \boldsymbol{x})\} = \delta^*(\cdot - \boldsymbol{x}) \in \mathcal{H}$ when indexed by $\boldsymbol{x}$ is called the reproducing kernel of the Hilbert space.

The "reproducing" qualifier refers to the basic property that

$$(r_{\mathcal{H}}(\cdot, \boldsymbol{x}), f)_{\mathcal{H}} = (\delta^*(\cdot - \boldsymbol{x}), f)_{\mathcal{H}} = \langle \delta(\cdot - \boldsymbol{x}), f \rangle_{\mathcal{H}' \times \mathcal{H}} = f(\boldsymbol{x})$$

for all $f \in \mathcal{H}$ and any $\boldsymbol{x} \in \mathbb{R}^d$. We now use Theorem 2 to obtain a multi-kernel extension of Schölkopf's celebrated representer theorem for learning in RKHS [8]. The classical theorem corresponds to the scenario of a single RKHS with $N = 1$, for which the underlying kernel is simply $r(\cdot, \boldsymbol{y}) = r_{\mathcal{H}}(\cdot, \boldsymbol{y}) = \delta^*(\cdot - \boldsymbol{y})$.

Corollary 1 *(Multi-kernel expansion in RKHS). Let us consider the following setting:*

- *A series of reproducing-kernel Hilbert spaces $\mathcal{H}_1, \dots, \mathcal{H}_N$ whose members are functions on $\mathbb{R}^d$. The reproducing kernel of $\mathcal{H}_n$ is $r_{\mathcal{H}_n} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.*
- *A strictly convex loss functional $E : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$.*

- *A strictly increasing and convex function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$.*
- *An absolute norm $\| \cdot \|_{\mathcal{Z}}$ on $\mathbb{R}^N$.*

Then, for any given series of points $(\boldsymbol{x}_m, y_m) \in \mathbb{R}^{d+1}$, $m = 1, \ldots, M$, the multi-component data-fitting problem

$$S = \arg \min_{f = f_1 + \cdots + f_N : f_n \in \mathcal{H}_n} \left(\sum_{m=1}^{M} E\big(y_m, f(\boldsymbol{x}_m)\big) + \psi\big(\big\|(\|f_1\|_{\mathcal{H}_1}, \ldots, \|f_N\|_{\mathcal{H}_N})\big\|_{\mathcal{Z}'}\big) \right) \tag{28}$$

always admits a global solution of the form

$$f_0(\boldsymbol{x}) = \sum_{m=1}^{M} a_m r(\boldsymbol{x}, \boldsymbol{x}_m) \tag{29}$$

with $(a_1, \ldots, a_M) \in \mathbb{R}^M$, where the underlying multi-kernel is given by

$$r(\boldsymbol{x}, \boldsymbol{y}) = \sum_{n=1}^{N} \alpha_n r_{\mathcal{H}_n}(\boldsymbol{x}, \boldsymbol{y}), \tag{30}$$

with suitable weights $(\alpha_1, \cdots, \alpha_N) \in \mathbb{R}^N_{\geq 0}$. Moreover, the solution is unique if $\psi(\cdot)$ and $\| \cdot \|_{\mathcal{Z}'}$ are both strictly convex.

Proof. Problem (28) is a special case of (27) with $\mathcal{X}'_n = \mathcal{H}_n = \mathcal{H}''_n$ and $\nu_m = \delta(\cdot - \boldsymbol{x}_m) \in \mathcal{X}_n = \mathcal{H}'_n$, due to the RKHS property. We then apply (22) in Theorem 2 with $\nu_{m,n} = \delta(\cdot - \boldsymbol{x}_m)$, which gives the parametric form of the solution components with $\varphi_{m,n} = r_{\mathcal{H}_n}(\cdot, \boldsymbol{x}_m) = \mathrm{J}_{\mathcal{H}'_n}\{\delta(\cdot - \boldsymbol{x}_m)\}$ where $r_{\mathcal{H}_n}$ is the reproducing kernel of $\mathcal{H}_n$. The summation constraint $f_0 = f_{0,1} + \cdots + f_{0,N}$ and the property that the coefficients $\boldsymbol{a} \in \mathbb{R}^M$ are shared by all components then gives (29). $\quad \square$

The interesting aspect in Corollary 1 is that the underlying kernel given by (30) is tunable, which offers flexibility and is in the line with certain forms of multiple-kernel learning [28].

The simplest choice of regularization functional in (28) is the weighted sum $\sum_{n=1}^{N} \lambda_n \|f_n\|^2_{\mathcal{H}_n}$, where the $\lambda_n > 0$ are appropriate regularization parameters. This corresponds to $\psi(\cdot) = |\cdot|^2$ in (28) with the outer weighted Euclidean norm

$$\|\boldsymbol{y}\|_{\mathcal{Z}'} = \|\boldsymbol{y}\|_{2,\boldsymbol{\lambda}} = \big(\lambda_1 y_1^2 + \cdots + \lambda_N y_N^2\big)^{\frac{1}{2}},$$

whose dual is $\| \cdot \|_{2,\boldsymbol{\mu}}$ with $\boldsymbol{\mu} = (1/\lambda_1, \ldots, 1/\lambda_N)$. Correspondingly, the ($\mathcal{Z}$-$\mathcal{Z}'$) conjugate of $\boldsymbol{y} \in \mathcal{Z}'$ is

$$\boldsymbol{y}^* = (\lambda_1 y_1, \ldots, \lambda_N y_N) \in \mathcal{Z}$$

which, with the help of (21) in Theorem 2, then yields $\alpha_n = y_n / y_n^* = 1/\lambda_n$ and offers direct control over (30).

By selecting the outer norm $\| \cdot \|_{\mathcal{Z}'}$ to be non-Euclidean, one can also make the kernel-shaping effect in (30) data-dependent, with the effect becoming more pronounced as we relax the "strictness" of the convexity requirement. The prototypical case of a regularization functional that falls into the latter category is $\lambda \sum_{n=1}^{N} \|f_n\|_{\mathcal{H}_n}$ (mixed ℓ_1-norm) with parameter $\lambda > 0$, which favors sparse kernel configurations. Micchelli and Pontil [29] have shown that the corresponding minimization problem is equivalent to a special instance of kernel learning where the "optimal" kernel $r(\cdot, \boldsymbol{x})$ is selected within the multi-kernel dictionary

$$\mathcal{K} = \left\{ \sum_{n=1}^{N} \alpha_n r_{\mathcal{H}_n}(\cdot, \boldsymbol{x}) : \alpha_n \geq 0, \alpha_1 + \cdots + \alpha_N = 1 \right\},$$

which is consistent with the last statement in Theorem 2.

The approach is generalizable to reflexive Banach spaces with the caveat that the resulting representer model is no longer linear.

Definition 6 *(See [54,55]).* A strictly convex and reflexive Banach space $\mathcal{B}$ of functions on $\mathbb{R}^d$ is called a reproducing-kernel Banach space (RKBS) if $\delta(\cdot - \boldsymbol{x}) \in \mathcal{B}'$ for any $\boldsymbol{x} \in \mathbb{R}^d$. Then, the unique representer $r_{\mathcal{B}}(\cdot, \boldsymbol{x}) = \mathsf{J}_{\mathcal{B}'}\{\delta(\cdot - \boldsymbol{x})\} = \delta^*(\cdot - \boldsymbol{x}) \in \mathcal{B}$ when indexed by $\boldsymbol{x}$ is called the reproducing kernel of the Banach space.

We can then consider the direct Banach counterpart of the multicomponent problem in Corollary 1 with $f_n \in \mathcal{B}_n = \mathcal{X}'_n$, $\delta(\cdot - \boldsymbol{x}_m) \in \mathcal{B}'_n = \mathcal{X}''_n = \mathcal{X}_n$ and derive the general parametric form of the solution as

$$f_0 = f_1 + \cdots + f_N \text{ with } f_n = \alpha_n \mathsf{J}_{\mathcal{B}'_n} \left\{ \sum_{m=1}^{M} a_m \delta(\cdot - \boldsymbol{x}_m) \right\} \tag{31}$$

with $(a_m) \in \mathbb{R}^M$ and $(\alpha_n) \in \mathbb{R}_{\geq 0}^N$. The result provided by (31) is new to the best of our knowledge. It is the multi-component extension of [55, Theorem 2], which is itself a reformulation and slight generalization of [54, Theorem 19].

4.2. Sparse signal representation in dictionaries

While the use of classical Hilbertian smoothness (e.g., Sobolev) norms lends itself to a closed-form resolution of (27), the underlying multicomponent model takes its full power when the regularization norms are not strictly convex and promote sparsity. This statement is supported by a large body of work in compressed sensing (CS). To show how this fits the present formulation, we now consider the prototypical CS problem: the recovery of a vector $\mathbf{x} \in \mathbb{R}^N$ from its linear measurements $\mathbf{H}\mathbf{x} = \mathbf{y} \in \mathbb{R}^M$, where $\mathbf{H}$ is the system matrix with M much smaller than N. Given a series of invertible matrices $\mathbf{L}_i \in \mathbb{R}^{N \times N}$, we then specify the component Banach spaces $\mathcal{X}'_i = (\mathbb{R}^N, \|\cdot\|_{\mathcal{X}'_i})$ with $\|\mathbf{x}\|_{\mathcal{X}'_i} = \|\mathbf{L}_i \mathbf{x}\|_1$. There, the use of the ℓ_1-norm is intended to promote sparsity in the transformed domain associated with $\mathbf{L}_i$.

In order to be able to apply Theorem 2, we need to identify the predual spaces $\mathcal{X}_i$ as well as the extremal points of the unit ball in $\mathcal{X}'_i$. To that end, we invoke Proposition 1 with the primary pair of dual spaces $\mathcal{X} = (\mathbb{R}^N, \|\cdot\|_\infty)$, $\mathcal{X}' = (\mathbb{R}^N, \|\cdot\|_1)$ and with $\mathsf{T}_i : \mathbf{x} \mapsto \mathbf{L}_i^T \mathbf{x}$. This then results in the identification of the predual Banach space $\mathcal{X}_i = \mathsf{T}_i(\mathcal{X}) = (\mathbb{R}^N, \|\cdot\|_{\mathcal{X}_i})$ with $\|\mathbf{x}\|_{\mathcal{X}_i} = \|\mathbf{L}_i^{-T}\mathbf{x}\|_\infty$ and its dual $\mathcal{X}'_i = (\mathbb{R}^N, \|\cdot\|_{\mathcal{X}'_i})$ with

$$\|\mathbf{y}\|_{\mathcal{X}'_i} = \sup_{\|\mathbf{x}\|_{\mathcal{X}_i}=1} \mathbf{y}^T \mathbf{x} = \sup_{\|\mathbf{L}_i^{-T}\mathbf{x}\|_\infty=1} \mathbf{y}^T \mathbf{x} = \sup_{\|\mathbf{v}\|_\infty=1} \mathbf{y}^T \mathbf{L}_i^T \mathbf{v} = \|\mathbf{L}_i \mathbf{y}\|_1, \tag{32}$$

which is consistent with the definition $\|\mathbf{x}\|_{\mathcal{X}'_i} = \|\mathbf{L}_i \mathbf{x}\|_1 = \|\mathbf{c}_i\|_1$.

It is well known that the extremal points of the unit ball in $(\mathbb{R}^N, \|\cdot\|_1)$ are $\pm \mathbf{e}_n$ where $\mathbf{e}_n$ is the nth element of the canonical basis with $[\mathbf{e}_n]_m = \delta_{m-n}$. We then apply the isometric isomorphism $\mathcal{Y}'_i = \mathsf{T}_i^{-1*}(\mathcal{X}')$ with $\mathsf{T}_i^{-1*} : \mathbf{e} \mapsto \mathbf{L}_i^{-1}\mathbf{e}$ (see Properties 2 and 4 in Proposition 1) to deduce that $\mathrm{Ext}(B_{\mathcal{X}'_i}) = \{\pm \mathbf{u}_{n,i}\}_{n=1}^N$, where $B_{\mathcal{X}'_i}$ is the unit ball of the Banach space $(\mathbb{R}^N, \|\cdot\|_{\mathcal{X}'_i})$ and

$$\mathbf{u}_{n,i} = \mathbf{L}_i^{-1}\mathbf{e}_n. \tag{33}$$

In fact, the latter defines a (sub)-dictionary whose elements are the row vectors of $\mathbf{L}_i^{-1}$.

We now have the tools to characterize the solution(s) of the multi-component ℓ_1 optimization problem

$$S = \arg \min_{\mathbf{x}=\mathbf{x}_1+\cdots+\mathbf{x}_I \in \mathbb{R}^N} \left(\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^I \|\mathbf{L}_i \mathbf{x}_i\|_1 \right), \tag{34}$$

which is a special case of (27) with $\nu_m : \mathbf{x} \mapsto \mathbf{h}_n^T \mathbf{x}$, $\psi = \mathrm{Id}$, and $\|\cdot\|_{\mathcal{Z}'} = \|\cdot\|_1$ where $\mathbf{h}_n^T$ is the nth row vector of $\mathbf{H}$. First, we confirm that the problem is well posed by observing that $\mathbf{h} \in \mathcal{X}_i$ for any measurement vector $\mathbf{h} \in \mathbb{R}^N$. In other words, (34) always admits a solution, albeit not necessarily a unique one. To obtain the parametric form of the extremal solution components $(\mathbf{x}_{0,1}, \ldots, \mathbf{x}_{0,M})$, we apply the last part of Theorem 2, which yields

$$\mathbf{x}_{0,i} = \sum_{k=1}^{K_i} c_{k,i} \mathbf{u}_{n_k,i}$$

with $\sum_i^I K_i \le M$ and $\sum_{i=1}^I \|\mathbf{L}_i \mathbf{x}_{0,i}\|_1 = \sum_{i=1}^I \|\mathbf{c}_i\|_1$. This may also be rewritten as

$$\mathbf{x}_0 = \sum_{i=1}^I \mathbf{x}_{0,i} = \sum_{k=1}^{K_0} c_k \mathbf{u}_{n_k, i_k} \tag{35}$$

with $K_0 \le M$ and some appropriate weights $(c_k) \in \mathbb{R}^{K_0}$ and $\sum_{i=1}^I \|\mathbf{x}_{i,0}\|_{\mathcal{X}_i'} = \|\mathbf{c}\|_1$. In effect, (35) tells us that the solution is constructed by picking K_0 atoms (with $K_0 \le M$) in an enlarged dictionary

$$\mathbf{U} = [\mathbf{u}_{1,1} \ \cdots \ \mathbf{u}_{N,1} \ \mathbf{u}_{1,2} \ \cdots \mathbf{u}_{2,N} \ \cdots \cdots \ \mathbf{u}_{I,N}] \in \mathbb{R}^{N \times (I \times N)},$$

which is formed from the unions of the $\mathbf{u}_{n,i}$ in (33) with $i = 1, \ldots, I$, $n = 1, \ldots, N$. The result in (35) also motivates us to reformulate Problem (34) in a more familiar "synthesis" form

$$S = \arg \min_{\mathbf{c} \in \mathbb{R}^{N \times I}} \left(\|\mathbf{y} - \mathbf{H}\mathbf{U}\mathbf{c}\|_2^2 + \lambda \|\mathbf{c}\|_1 \right), \tag{36}$$

where $\mathbf{c} \in \mathbb{R}^{N \times I}$ is an augmented parameter vector. The latter is the standard LASSO formulation for the recovery of a signal subject to the constraint that it has a sparse representation in some predefined dictionary $\mathbf{U}$. The new elements here are the link with the "analysis" form (34) and the guarantee of the existence of a "sparse" solution with $K_0 \le M$, irrespective of whether the conditions for uniqueness (e.g., restricted isometry) of CS are met or not.

4.3. Signal recovery problems involving mixed norms

While the examples of Section 4.1 and 4.2 are fairly classical, we can use our high-level results to derive some new represeter theorems, such as the following example which involves a combination of smoothness (RKHS) and sparsity-promoting regularizations.

As prerequisite, we need to specify dual pairs of Banach spaces that are matched to specific norms and regularization operators L_i. To that end, we assume that the linear operator $\mathrm{T} = \mathrm{L}^*$ is injective on some primary space $\mathcal{X}$ and recall the relevant results from Proposition 1:

1. The operator L^* is invertible on its range $\mathrm{L}^*(\mathcal{X})$.
2. The dual pair of Banach spaces $\mathcal{B} = \mathrm{L}^*(\mathcal{X})$ and $\mathcal{B}' = \mathrm{L}^{-1}(\mathcal{X}')$ is isometrically isomorphic to $(\mathcal{X}, \mathcal{X}')$.

3. The operator L isometrically maps $\mathcal{B}' = \mathrm{L}^{-1}(\mathcal{X}')$ to $\mathcal{X}'$.
4. If $\mathcal{X} = L_2(\mathbb{R}^d)$, then $\mathcal{H} = \mathrm{L}^*(L_2(\mathbb{R}^d))$ and $\mathcal{H}' = \mathrm{L}^{-1}(L_2(\mathbb{R}^d))$ are both Hilbert spaces, while the underlying Riesz map is $\mathrm{J}_{\mathcal{H}} = \mathrm{L}^{-1}\mathrm{L}^{-1*} : \mathcal{H} \to \mathcal{H}'$.

When we choose $\mathcal{X} = L_2(\mathbb{R}^d)$, the space $\mathcal{X}$ is its own dual and is usually associated with Tikhonov regularization. The other fundamental scenario is $(\mathcal{X}, \mathcal{X}') = (C_0(\mathbb{R}^d), \mathcal{M}(\mathbb{R}^d))$. There, $C_0(\mathbb{R}^d)$ is the space of continuous functions that decay at infinity, while its dual $\mathcal{M}(\mathbb{R}^d) = (C_0(\mathbb{R}^d))'$ is the space of bounded Radon measures on $\mathbb{R}^d$ [44]. The regularization norm $\|\cdot\|_{\mathcal{M}}$, also known as "total variation" in the sense of measure theory, is often used in applications since it promotes sparsity [56,57,17].

Corollary 2 (*Representer theorem for mixed-norm regularization*). *Let us consider the following setting:*

- *An operator L_1^* that is injective on $C_0(\mathbb{R}^d)$ and the corresponding dual pair of Banach spaces $\mathcal{B}_1 = \mathrm{L}_1^*(\mathcal{C}_0(\mathbb{R}^d))$ and $\mathcal{B}_1' = \mathrm{L}_1^{-1}(\mathcal{M}(\mathbb{R}^d))$.*
- *An operator L_2^* that is injective on $L_2(\mathbb{R}^d)$ and the corresponding dual pair of Hilbert spaces $\mathcal{H}_2 = \mathrm{L}_2^*(L_2(\mathbb{R}^d))$ and $\mathcal{H}_2' = \mathrm{L}_2^{-1}(L_2(\mathbb{R}^d))$.*
- *The linear measurement operator $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_M) : \mathcal{B}_1' + \mathcal{H}_2' \to \mathbb{R}^M$ with $\nu_1, \ldots, \nu_M \in \mathcal{B}_1 \cap \mathcal{H}_2$.*
- *A strictly convex loss functional $E : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}^+$.*
- *Two adjustable weights $\lambda_1, \lambda_2 \in \mathbb{R}^+$.*

Then, for any given $\boldsymbol{y} \in \mathbb{R}^M$, the two-component regularized inverse problem

$$S = \arg\min_{f = f_1 + f_2 : (f_1, f_2) \in \mathcal{B}_1' \times \mathcal{H}_2'} \left(E(\boldsymbol{y}, \boldsymbol{\nu}(f)) + \lambda_1 \|\mathrm{L}_1 f_1\|_{\mathcal{M}}^2 + \lambda_2 \|\mathrm{L}_2 f_2\|_{L_2}^2 \right) \tag{37}$$

has a nonempty (and weak*-compact) solution set S. Any solution can be written as $f_0 = f_{0,1} + f_{0,2}$ with $f_{0,1} \in S|_{\mathcal{B}_1}$ and a unique

$$f_{0,2} = \sum_{m=1}^{M} \widetilde{a}_m \varphi_m, \tag{38}$$

with $\varphi_m = \mathrm{L}_2^{-1}\mathrm{L}_2^{-1*}\{\nu_m\}$ and $(\widetilde{a}_m) \in \mathbb{R}^M$, that is common to all solutions. Moreover, the extremal points of $S|_{\mathcal{B}_1}$ can all be expressed as

$$f_{0,1} = \sum_{k=1}^{K_1} c_k h_1(\cdot, \boldsymbol{\tau}_k) \tag{39}$$

with $K_1 \leq M$, $(c_k) \in \mathbb{R}^{K_1}$ and $h_1(\cdot, \boldsymbol{\tau}_k) = \mathrm{L}_1^{-1}\{\delta(\cdot - \boldsymbol{\tau}_k)\}$, where $\boldsymbol{\tau}_1, \ldots, \boldsymbol{\tau}_{K_1} \in \mathbb{R}^d$ are adaptive centers.

Proof. Problem (37) is a special case of (27) with $N = 2$, $\mathcal{X}_1' = \mathcal{B}_1'$, $\mathcal{X}_2' = \mathcal{H}_2'$, $\psi(\cdot) = |\cdot|^2$ (strictly convex), and $\|(x_1, x_2)\|_{\mathcal{Z}'} = \sqrt{\lambda_1 x_1^2 + \lambda_2 x_2^2}$. The hypotheses are such that the conditions for the validity of Theorem 2 are met. The parametric form of the second (unique) solution component is then given by (22) with $\widetilde{a}_m = \alpha_2 a_m$, $\nu_{m,n} = \nu_m$ and $\mathrm{J}_{\mathcal{H}_2} = \mathrm{L}_2^{-1}\mathrm{L}_2^{-1*}$. To characterize the extremal points of the first component, we use (24). To identify the relevant atoms, we recall that the extremal points of the unit ball in $\mathcal{M}(\mathbb{R}^d)$ are given by $\{\pm\delta(\cdot - \boldsymbol{\tau})\}_{\boldsymbol{\tau} \in \mathbb{R}^d}$ [22]. We then make use of the fourth property in Proposition 1 to obtain the form of an extremal point of the unit ball in $\mathcal{B}_1'$: $e_k = \pm\mathrm{L}_1^{-1}\{\delta(\cdot - \boldsymbol{\tau}_k)\}$ with shift parameter $\boldsymbol{\tau}_k \in \mathbb{R}^d$. $\quad\square$

As slight variant of (37), we may consider an outer ℓ_1 norm with $\psi(t) = t$, which yields a regularization of the form $\lambda_1 \|\mathrm{L}_1 f_1\|_{\mathcal{M}} + \lambda_2 \|\mathrm{L}_2 f_2\|_{L_2}$. The resulting solution takes the same functional form, with the caveat that there may no longer exist a single Hilbert-space component $f_{0,2}$ that would be common to all solutions.

5. Minimization of semi-norms

We now consider the scenario of a native Banach space $\mathcal{X}'$ that has a direct-sum decomposition as $\mathcal{X}' = \mathcal{U}' \oplus \mathcal{N}_{\boldsymbol{p}}$, where $\mathcal{U}'$ is the dual of some primary Banach space $(\mathcal{U}, \|\cdot\|_{\mathcal{U}})$ and where the complementary space $\mathcal{N}_{\boldsymbol{p}}$ is spanned by the finite-dimensional basis $\boldsymbol{p} = (p_1, \ldots, p_{N_0})$. Since $\mathcal{N}_{\boldsymbol{p}} = \operatorname{span}\{p_n\}_{n=1}^{N_0}$ is of dimension N_0 and hence also reflexive, the same holds true for its continuous dual $\mathcal{N}_{\boldsymbol{p}}'$. Moreover, due to the direct-sum property, there exists a unique biorthonormal set of generators $p_1^*, \ldots, p_{N_0}^* \in \mathcal{X}$ such that $\mathcal{N}_{\boldsymbol{p}}' = \operatorname{span}\{p_n^*\}_{n=1}^{N_0} = \mathcal{N}_{\boldsymbol{p}^*}$ and

$$\langle p_m^*, p_n \rangle = \delta_{m,n}$$
$$\langle p_n^*, s \rangle = 0$$

for any $s \in \mathcal{U}'$ and $m, n \in \{1, \ldots, N_0\}$. This allows us to specify the canonical projector $\operatorname{Proj}_{\mathcal{N}_{\boldsymbol{p}}} : \mathcal{X}' \to \mathcal{N}_{\boldsymbol{p}}$ as

$$\operatorname{Proj}_{\mathcal{N}_{\boldsymbol{p}}}\{f\} = \sum_{n=1}^{N_0} \langle p_n^*, f \rangle p_n \tag{40}$$

for any $f \in \mathcal{X}'$. This identification also yields the complementary projector $\operatorname{Proj}_{\mathcal{U}'} : \mathcal{X}' \to \mathcal{U}'$ as $\operatorname{Proj}_{\mathcal{U}'} = (\mathrm{Id} - \operatorname{Proj}_{\mathcal{N}_{\boldsymbol{p}}})$. Likewise, by interchanging the role of the synthesis and analysis functionals in (40), we identify the canonical projector $\operatorname{Proj}_{\mathcal{N}_{\boldsymbol{p}^*}} : \mathcal{X} \to \mathcal{N}_{\boldsymbol{p}^*}$ as

$$\operatorname{Proj}_{\mathcal{N}_{\boldsymbol{p}^*}}\{\nu\} = \sum_{n=1}^{N_0} \langle p_n, \nu \rangle p_n^*$$

for any $\nu \in \mathcal{X}$. We now have the means to specify and bound the norm of any $f \in \mathcal{X}'$ as

$$\|f\|_{\mathcal{X}'} = (\|f\|_{\mathcal{U}'}, \|\operatorname{Proj}_{\mathcal{N}_{\boldsymbol{p}}}\{f\}\|_{\mathcal{N}_p})_{\mathcal{Z}'} \le \|f\|_{\mathcal{U}'} + \|\operatorname{Proj}_{\mathcal{N}_{\boldsymbol{p}}}\{f\}\|_{\mathcal{N}_p}$$

where the functional $f \mapsto \|f\|_{\mathcal{U}'} \stackrel{\triangle}{=} \|\operatorname{Proj}_{\mathcal{U}'} f\|_{\mathcal{U}'}$ is a semi-norm (resp., a norm) over $\mathcal{X}'$ (resp., $\mathcal{U}'$). Because the p_n are linearly independent, we also note that the standard description of $\mathcal{U}'$ as the complement of $\mathcal{N}_{\boldsymbol{p}}$ in $\mathcal{X}'$, given by

$$\mathcal{U}' = \{s \in \mathcal{X}' : \operatorname{Proj}_{\mathcal{N}_{\boldsymbol{p}}}\{s\} = \sum_{n=1}^{N_0} \langle p_n^*, s \rangle p_n = 0\},$$

is equivalent to

$$\mathcal{U}' = \{s \in \mathcal{X}' : \boldsymbol{p}^*(s) = \boldsymbol{0}\}$$

where $\boldsymbol{p}^*(f) = (\langle p_1^*, f \rangle, \ldots, \langle p_{N_0}^*, f \rangle)$ is a vector-valued functional $\mathcal{X}' \to \mathbb{R}^{N_0}$. Likewise, we have that

$$\mathcal{U} = \{u \in \mathcal{X} : \boldsymbol{p}(u) = \boldsymbol{0}\},$$

which, once again, capitalizes on the biorthonormality of $(\boldsymbol{p}^*, \boldsymbol{p})$.

The consideration of such a direct-sum decomposition of a Banach space $\mathcal{X}'$ is relevant to inverse problems because it suggests that one can substitute the original regularization term $\|f\|_{\mathcal{X}'}$ by the semi-norm $\|f\|_{\mathcal{U}'}$ when one wants to favor solutions with a strong contribution in $\mathcal{N}_{\boldsymbol{p}}$, the null space of the semi-norm. This is a standard technique in spline theory, albeit within the classical context of RKHS spaces with $\|f\|_{\mathcal{U}'} = \|\mathrm{L}f\|_{L_2}$, where L is a suitable differential operator (e.g., a higher-order derivative or fractional Laplacian) with a null space $\mathcal{N}_{\boldsymbol{p}}$ that consists of polynomials of degree n [31–33]. We now show how this technique can be extended in full generality to Banach spaces. The basic requirement for this extension is that the inverse problem be well-posed over $\mathcal{N}_{\boldsymbol{p}}$. This is made explicit in (54), which is equivalent to the fourth condition in Theorem 3.

Theorem 3 *(General representer theorem for Banach semi-norms). Let us consider the following setting:*

- *A dual pair of Banach spaces $(\mathcal{X} = \mathcal{U} \oplus \mathcal{N}_{\boldsymbol{p}^*}, \mathcal{X}' = \mathcal{U}' \oplus \mathcal{N}_{\boldsymbol{p}})$, where $\mathcal{N}_{\boldsymbol{p}} = \mathcal{N}'_{\boldsymbol{p}^*}$ is the vector space spanned by the finite-dimensional basis $\boldsymbol{p} = (p_1, \ldots, p_{N_0})$.*
- *The analysis subspace $\mathcal{N}_{\boldsymbol{\nu}} = \mathrm{span}\{\nu_m\}_{m=1}^M \subset \mathcal{X}$, with the ν_m being linearly independent and $M > N_0$.*
- *The linear measurement operator $\boldsymbol{\nu} : \mathcal{X}' \to \mathbb{R}^M : f \mapsto (\langle \nu_1, f \rangle, \ldots, \langle \nu_M, f \rangle)$.*
- *The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{N_0} \in \mathbb{R}^M$ with $[\mathbf{v}_n]_m = \langle \nu_m, p_n \rangle$ are linearly independent; they admit a complementary set $\{\mathbf{u}_1, \ldots, \mathbf{u}_{M-N_0}\}$ in $\mathbb{R}^M$ such that $\mathbb{R}^M = \mathrm{span}\{\mathbf{v}_n\}_{n=1}^{N_0} \oplus \mathrm{span}\{\mathbf{u}_m\}_{m=1}^{M-N_0}$.*
- *A proper, lower-semicontinuous, coercive, and convex loss functional $E : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}^+ \cup \{+\infty\}$.*
- *Some arbitrary strictly increasing and convex function $\psi : \mathbb{R}^+ \to \mathbb{R}^+$.*

Then, for any fixed $\boldsymbol{y} \in \mathbb{R}^M$, the solution set of the generic optimization problem

$$S = \arg \min_{f \in \mathcal{X}'} \left(E(\boldsymbol{y}, \boldsymbol{\nu}(f)) + \psi(\|f\|_{\mathcal{U}'}) \right) \tag{41}$$

is nonempty, convex, and weak*-compact.

If E is strictly convex, or if it imposes the equality constraint $\boldsymbol{y} = \boldsymbol{\nu}(f)$, then any solution $f_0 \in S \subset \mathcal{X}'$ has a unique decomposition as $f_0 = p_0 + s_0$ with $p_0 \in \mathcal{N}_{\boldsymbol{p}}$ and $s_0 \in \mathcal{U}'$ the $(\mathcal{U}', \mathcal{U})$-conjugate of a common $\widetilde{\nu}_0 \in \mathcal{U}$ whose generic form is

$$\widetilde{\nu}_0 = \sum_{m=1}^{M-N_0} a_m \widetilde{\nu}_m \in \mathcal{N}_{\boldsymbol{\nu}} \cap \mathcal{U} \tag{42}$$

with suitable coefficients $\boldsymbol{a} = (a_m) \in \mathbb{R}^{M-N_0}$ and reduced basis functions $\widetilde{\nu}_m = \widetilde{\mathbf{u}}_m^T \boldsymbol{\nu} \in \mathcal{U}$, where $\widetilde{\mathbf{u}}_m \in \mathbb{R}^M$ is the unique (biorthogonal) vector such that $\widetilde{\mathbf{u}}_m^T \mathbf{v}_n = 0$ and $\widetilde{\mathbf{u}}_m^T \mathbf{u}_{m'} = \delta_{m,m'}$ for any $m, m' \in \{1, \ldots, M-N_0\}$ and $n \in \{1, \ldots, N_0\}$.

Depending on the Banach characteristics of $\mathcal{U}'$, this then results in the following explicit description of the solution(s):

- *If $\mathcal{U}'$ is a Hilbert space and ψ is strictly convex, then the solution f_0 is unique and admits the linear representation*

$$f_0 = \sum_{m=1}^{M-N_0} a_m \varphi_m + \sum_{n=1}^{N_0} b_n p_n, \tag{43}$$

with coefficients $(\boldsymbol{a}, \boldsymbol{b}) \in \mathbb{R}^M$ and basis functions $p_n \in \mathcal{N}_{\boldsymbol{p}}$, $\varphi_m = \mathrm{J}_{\mathcal{U}}\{\widetilde{\nu}_m\} \in \mathcal{U}'$, where $\mathrm{J}_{\mathcal{U}}$ is the Riesz map $\mathcal{U} \to \mathcal{U}'$.

- If $\mathcal{U}'$ is a strictly convex Banach space and ψ is strictly increasing, then the solution is unique and admits the parametric representation

$$f_0 = \mathrm{J}_{\mathcal{U}} \left\{ \sum_{m=1}^{M-N_0} a_m \widetilde{\nu}_m \right\} + \sum_{n=1}^{N_0} b_n p_n \tag{44}$$

  where $\mathrm{J}_{\mathcal{U}}$ is the (nonlinear) duality operator $\mathcal{U} \to \mathcal{U}'$.
- Otherwise, when $\mathcal{U}'$ is not strictly convex, the solution set is the weak*-closure of the convex hull of its extremal points, which can all be expressed as

$$f_0 = \sum_{k=1}^{K_0} c_k e_k + \sum_{n=1}^{N_0} b_n p_n \tag{45}$$

  for some $K_0 \leq (M - N_0)$, $c_1, \ldots, c_{K_0} \in \mathbb{R}$, where $e_1, \ldots, e_{K_0} \in \mathcal{U}'$ are some extremal points of the unit ball $B_{\mathcal{U}'} = \{s \in \mathcal{U} : \|s\|_{\mathcal{U}'} \leq 1\}$. The vector $\boldsymbol{b} = (b_n) \in \mathbb{R}^{N_0}$ that characterizes the null-space component of f_0 is unique and common to all solutions whenever E is strictly convex and $\mathcal{N}_{\boldsymbol{p}^*} \subset \mathcal{N}_{\boldsymbol{\nu}}$.

Before proceeding with the proof of Theorem 3, we detail the way in which the reduced basis $\widetilde{\boldsymbol{\nu}} = (\widetilde{\nu}_1, \ldots, \widetilde{\nu}_{M-N_0})$ in (42) is constructed. To that end, we first define the cross-correlation matrix $\mathbf{V} = [\mathbf{v}_1 \; \cdots \; \mathbf{v}_{N_0}] = \boldsymbol{\nu}(\boldsymbol{p}) \in \mathbb{R}^{M \times N_0}$ with $[\mathbf{V}]_{m,n} = \langle \nu_m, p_n \rangle$.

Proposition 2 (Direct-sum decomposition of the measurement space). Let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_M) \in \mathcal{X}^M$ with $\mathcal{X} = \mathcal{U} \oplus \mathcal{N}_{\boldsymbol{p}^*}$ and $\boldsymbol{p} = (p_1, \ldots, p_{N_0}) \in (\mathcal{X}')^{N_0}$ be two vectors of linear functionals such that the matrix $\mathbf{V} = \boldsymbol{\nu}(\boldsymbol{p}) \in \mathbb{R}^{M \times N_0}$ is of rank N_0. Then, one can always find three matrices $\mathbf{U} \in \mathbb{R}^{M \times (M-N_0)}$, $\widetilde{\mathbf{V}} \in \mathbb{R}^{M \times N_0}$, and $\widetilde{\mathbf{U}} \in \mathbb{R}^{M \times (M-N_0)}$ such that

$$\begin{bmatrix} \widetilde{\mathbf{U}}^T \\ \widetilde{\mathbf{V}}^T \end{bmatrix} \begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix} = \mathbf{I}_M. \tag{46}$$

Based on these matrices, $\boldsymbol{\nu} \in \mathcal{X}^M$ has a unique and reversible decomposition as

$$\boldsymbol{\nu} = \mathbf{U}\widetilde{\boldsymbol{\nu}} + \mathbf{V}\widetilde{\boldsymbol{p}}^*, \tag{47}$$

where

$$\widetilde{\boldsymbol{\nu}} = (\widetilde{\nu}_1, \ldots, \widetilde{\nu}_{M-N_0}) = \widetilde{\mathbf{U}}^T \boldsymbol{\nu} \in \mathcal{U}^{M-N_0} \tag{48}$$

$$\widetilde{\boldsymbol{p}}^* = (\widetilde{p}_1^*, \ldots, \widetilde{p}_{N_0}^*) = \widetilde{\mathbf{V}}^T \boldsymbol{\nu} \in \mathcal{X}^{N_0}. \tag{49}$$

In effect, this yields a decomposition of the measurement space $\mathcal{N}_{\boldsymbol{\nu}} = \mathrm{span}\{\nu_m\}_{m=1}^M$ as $\mathcal{N}_{\boldsymbol{\nu}} = \mathcal{N}_{\widetilde{\boldsymbol{p}}^*} \oplus \mathcal{N}_{\widetilde{\boldsymbol{\nu}}}$ with $\mathcal{N}_{\widetilde{\boldsymbol{\nu}}} = \mathrm{span}\{\widetilde{\nu}_m\}_{m=1}^{M-N_0} \subset \mathcal{U}$. In particular, if $\mathcal{N}_{\boldsymbol{p}^*} \subset \mathcal{N}_{\boldsymbol{\nu}}$, then there is a unique matrix $\widetilde{\mathbf{V}} \in \mathbb{R}^{M \times N_0}$ of rank N_0 such that $\widetilde{\mathbf{V}}^T \boldsymbol{\nu} = \boldsymbol{p}^*$ and such that the decomposition still applies with $\mathcal{N}_{\boldsymbol{\nu}} = \mathcal{N}_{\boldsymbol{p}^*} \oplus \mathcal{N}_{\widetilde{\boldsymbol{\nu}}}$.

Proof. Since the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_{N_0} \in \mathbb{R}^M$ are linearly independent, they can always be completed by adding some vectors $\mathbf{v}_{N_0+1} = \mathbf{u}_1, \ldots, \mathbf{v}_M = \mathbf{u}_{M-N_0}$ to form a basis of $\mathbb{R}^M$. The linear independence of the resulting family (basis property) is equivalent to the existence of a unique dual basis $\widetilde{\mathbf{v}}_1, \ldots, \widetilde{\mathbf{v}}_M \in \mathbb{R}^M$ such that

$$\langle \widetilde{\mathbf{v}}_m, \mathbf{v}_n \rangle = \widetilde{\mathbf{v}}_m^T \mathbf{v}_n = \delta_{m,n} \tag{50}$$

for $m, n \in \{1, \ldots, M\}$ (biorthonormality property). This means that any vector $\boldsymbol{y} \in \mathbb{R}^M$ has a unique (and reversible) decomposition as $\boldsymbol{y} = \sum_{m=1}^{M} \langle \widetilde{\mathbf{v}}_m, \boldsymbol{y} \rangle \mathbf{v}_m$. By collecting the expansion coefficients in the two vectors $\widetilde{\boldsymbol{b}} = (\langle \widetilde{\mathbf{v}}_1, \boldsymbol{y} \rangle, \ldots, \langle \widetilde{\mathbf{v}}_{N_0}, \boldsymbol{y} \rangle)$ and $\widetilde{\boldsymbol{y}} = (\langle \widetilde{\mathbf{v}}_{N_0+1}, \boldsymbol{y} \rangle, \ldots, \langle \widetilde{\mathbf{v}}_M, \boldsymbol{y} \rangle)$ and identifying the matrices $\widetilde{\mathbf{V}} = [\widetilde{\mathbf{v}}_1 \ \cdots \ \widetilde{\mathbf{v}}_{N_0}]$, $\mathbf{U} = [\mathbf{v}_{N_0+1} \ \cdots \ \mathbf{v}_M]$, and $\widetilde{\mathbf{U}} = [\widetilde{\mathbf{v}}_{N_0+1} \ \cdots \ \widetilde{\mathbf{v}}_M]$, we then observe that the decomposability of $\boldsymbol{y} \in \mathbb{R}^M$ is equivalent to

$$\boldsymbol{y} = \begin{bmatrix} \mathbf{U} & \mathbf{V} \end{bmatrix} \begin{bmatrix} \widetilde{\boldsymbol{y}} \\ \widetilde{\boldsymbol{b}} \end{bmatrix} \tag{51}$$

with

$$\widetilde{\boldsymbol{y}} = \widetilde{\mathbf{U}}^T \boldsymbol{y} \in \mathbb{R}^{M-N_0}, \quad \widetilde{\boldsymbol{b}} = \widetilde{\mathbf{V}}^T \boldsymbol{y} \in \mathbb{R}^{N_0}. \tag{52}$$

Likewise, the enabling biorthonormality property (50) is equivalent to the invertibility condition (46).

By substituting $\boldsymbol{y} \in \mathbb{R}^M$, $\widetilde{\boldsymbol{y}} \in \mathbb{R}^{M-N_0}$, and $\widetilde{\boldsymbol{b}} \in \mathbb{R}^{N_0}$ by $\boldsymbol{\nu}(f), \widetilde{\boldsymbol{\nu}}(f)$, and $\widetilde{\boldsymbol{p}}^*(f)$, respectively, we then rephrase (51) and (52) in term of functionals, which yields the reversible decomposition described by (47), (48), and (49). To prove that $\mathcal{N}_{\widetilde{\boldsymbol{\nu}}} \subset \mathcal{U} = \{u \in \mathcal{X} : \boldsymbol{p}(u) = 0\}$, we invoke the invertibility condition (46), which yields $\boldsymbol{p}(\widetilde{\boldsymbol{\nu}}) = \left(\widetilde{\boldsymbol{\nu}}(\boldsymbol{p}) \right)^T = (\widetilde{\mathbf{U}}^T \mathbf{V})^T = \mathbf{0}^T$. Since, for any $\boldsymbol{a} \in \mathbb{R}^M$, we have that $\boldsymbol{a}^T \mathbf{U} \widetilde{\boldsymbol{\nu}} \in \mathrm{span}\{\widetilde{\nu}_n\}_{n=1}^{M-N_0}$ and $\boldsymbol{a}^T \mathbf{V} \widetilde{\boldsymbol{p}}^* \in \mathrm{span}\{\widetilde{p}_n^*\}_{n=1}^{N_0}$ with the linear expansion of $\boldsymbol{a}^T \boldsymbol{\nu} \in \mathcal{N}_{\boldsymbol{\nu}}$ in the corresponding basis being reversible, we can interpret (47) as the direct-sum decomposition $\mathcal{N}_{\boldsymbol{\nu}} = \mathcal{N}_{\widetilde{\boldsymbol{\nu}}} \oplus \mathcal{N}_{\widetilde{\boldsymbol{p}}^*}$.

The inclusion $\mathcal{N}_{\boldsymbol{p}^*} \subset \mathcal{N}_{\boldsymbol{\nu}}$, together with the linear independence of the ν_m, is equivalent to the existence of a unique transformation matrix $\widetilde{\mathbf{V}}^T$ of rank N_0 such that $\boldsymbol{p}^* = \widetilde{\mathbf{V}}^T \boldsymbol{\nu}$. While this sets the matrix $\widetilde{\mathbf{V}} \in \mathbb{R}^{M \times N_0}$, one is still left with sufficiently many degrees of freedom to select the complementary matrices $\mathbf{U}$ and $\widetilde{\mathbf{U}}$ such that (46) holds. $\square$

We note that, irrespective of whether we fix $\widetilde{\mathbf{V}}$ (second part of Proposition 2) or not, there are generally infinitely many admissible choices for $\mathbf{U}$ in (46) and, hence, for the construction of the reduced basis $\widetilde{\boldsymbol{\nu}}$ defined by (48). This does not contradict the unicity of (42). Indeed, different choices of extension correspond to different biorthogonal bases $\widetilde{\boldsymbol{\nu}} = (\widetilde{\nu}_1, \ldots, \widetilde{\nu}_{M-N_0})$ of the same subspace.

5.1. Proof of Theorem 3

Proof. *(i) Existence*: The classical conditions that ensure the existence of a minimizer of the functional $F(f) = E\big(\boldsymbol{y}, \boldsymbol{\nu}(f)\big) + \psi\left(\|f\|_{\mathcal{U}'}\right)$ are that $F(f)$ should be proper, convex, (weak*-)lower-semi-continuous, and coercive over $\mathcal{X}'$. These higher-level properties also imply that the solution set S is convex, and weak*-compact.

The first three conditions follow from the listed assumptions and the general properties of a (semi-)norm—see argumentation in the proof of Theorem 1 in [42]. To establish coercivity, we recall that the hypothesis $\nu_m \in \mathcal{X}$ implies the continuity of $\nu_m : \mathcal{X}' \to \mathbb{R}$ due to the continuous embedding of $\mathcal{X}$ in its bidual $\mathcal{X}'' = (\mathcal{X}')'$. Consequently, there exists some constant $A > 0$ such that

$$\|\boldsymbol{\nu}(f)\|_2 \leq A \|f\|_{\mathcal{X}'} \tag{53}$$

for all $f \in \mathcal{X}'$. Likewise, the linear independence of the $\mathbf{v}_n$ and the property that all finite-dimensional norms are equivalent implies the existence of $B > 0$ such that, for any $p \in \mathcal{N}_{\boldsymbol{p}}$,

$$B \|p\|_{\mathcal{N}_{\boldsymbol{p}}} \leq \|\boldsymbol{\nu}(p)\|_2. \tag{54}$$

By using the direct-sum decomposition $f = s + p$ with $(s, p) \in \mathcal{U}' \times \mathcal{N}_{\boldsymbol{p}}$, $\|f\|_{\mathcal{X}'} \leq \|s\|_{\mathcal{U}'} + \|p\|_{\mathcal{N}_{\boldsymbol{p}}}$, and $\|s\|_{\mathcal{U}'} = \|f\|_{\mathcal{U}'}$, we readily deduce that

$$\|\boldsymbol{\nu}(f)\|_2 \geq \|\boldsymbol{\nu}(p)\|_2 - \|\boldsymbol{\nu}(s)\|_2 \geq B\|p\|_{\mathcal{N}_{\boldsymbol{p}}} - A\|s\|_{\mathcal{U}'} \geq B\|f\|_{\mathcal{X}'} - (A + B)\|f\|_{\mathcal{U}'}, \qquad (55)$$

where we have made use of the triangle inequality and the two previous bounds. Let us now consider some sequence (f_m) in $\mathcal{X}'$ with $f_m = (s_m, q_m) \in \mathcal{U}' \times \mathcal{N}_{\boldsymbol{p}}$ such that $\|f_m\|_{\mathcal{X}'} \geq \|f_n\|_{\mathcal{X}'}$ for $m \geq n$ and $\lim_{m \to \infty} \|f_m\|_{\mathcal{X}'} = \infty$. Then, there are two possible asymptotic behaviors for the norm of $s_m = \mathrm{Proj}_{\mathcal{U}'} f_m$:

1. The quantity $\|s_m\|_{\mathcal{X}'} = \|f_m\|_{\mathcal{U}'} \to \infty$ as $m \to \infty$, in which case $J(f_m) \to \infty$ due to the unboundedness of $\psi : \mathbb{R}^+ \to \mathbb{R}^+$ at infinity.
2. There exists a constant C such that $\|f_m\|_{\mathcal{U}'} \leq C$ for all m. By invoking (55), we get that $\|\boldsymbol{\nu}(f_m)\|_2 \to \infty$ as $m \to \infty$, which, in turn, gives $F(f_m) \to \infty$, due to the coercivity of $E(\cdot, \boldsymbol{y})$.

In summary, $F(f) \to \infty$ as $\|f\|_{\mathcal{X}'} \to \infty$, which is the required coercivity property.

(ii) Representation of a solution: The underlying direct-sum property implies that any $f_0 \in S \subset \mathcal{X}'$ has a unique decomposition as $f_0 = s_0 + p_0$ with $(s_0, p_0) \in \mathcal{U}' \times \mathcal{N}_{\boldsymbol{p}}$. To derive the parametric form of a solution, we momentarily assume that p_0 (and, hence, $\boldsymbol{\nu}(p_0) \in \mathbb{R}^M$) and $\boldsymbol{y}_0 = \boldsymbol{\nu}(f_0) \in \mathbb{R}^M$ are known. By making use of the decomposition of the measurement space in Proposition 2, we observe that the penalized component $s_0 \in \mathcal{U}'$ solves the equivalent constrained-optimization problem

$$s_0 \in S_{p_0, \boldsymbol{y}_0} = \arg \min_{s \in \mathcal{U}'} \|s\|_{\mathcal{U}'} \text{ s.t. } \boldsymbol{y}_0 - \boldsymbol{\nu}(p_0) = \boldsymbol{\nu}(s) = \mathbf{U}\widetilde{\boldsymbol{\nu}}(s) + \mathbf{V}\widetilde{\boldsymbol{p}}^*(s), \qquad (56)$$

where $\widetilde{\boldsymbol{\nu}} = \widetilde{\mathbf{U}}^T \boldsymbol{\nu} \in \mathcal{U}^{M-N_0}$ and $\widetilde{\boldsymbol{p}}^* = \widetilde{\mathbf{V}}^T \boldsymbol{\nu} \in \mathcal{X}^{N_0}$. We now show that the effective number of linear constraints in (56) is actually $(M - N_0)$ and not M, as may be thought on first inspection. To that end, we multiply the linear constraint by $\widetilde{\mathbf{U}}^T \in \mathbb{R}^{(M-N_0) \times M}$ on both sides and use the properties that $\widetilde{\mathbf{U}}^T \mathbf{U} = \mathbf{I}_{M-N_0}$ and $\widetilde{\mathbf{U}}^T \mathbf{V} = \mathbf{0}$ (see (46) in Proposition 2). This yields

$$s_0 \in S_{p_0, \boldsymbol{y}_0} = \arg \min_{s \in \mathcal{U}'} \|s\|_{\mathcal{U}'} \text{ s.t. } \widetilde{\boldsymbol{\nu}}(s) = \widetilde{\boldsymbol{y}}_0, \qquad (57)$$

where $\widetilde{\boldsymbol{y}}_0 = \widetilde{\mathbf{U}}^T \big(\boldsymbol{y}_0 - \boldsymbol{\nu}(p_0) \big) = \widetilde{\mathbf{U}}^T \boldsymbol{y}_0 \in \mathbb{R}^{M-N_0}$. This latter simplification occurs because $\widetilde{\mathbf{U}}^T \boldsymbol{\nu}(p) = \widetilde{\boldsymbol{\nu}}(p) = 0$ for all $p \in \mathcal{N}_{\boldsymbol{p}}$ by construction. The theoretical significance of the cancellation of $\widetilde{\mathbf{U}}^T \boldsymbol{\nu}(p_0)$ is that the above manipulation does not depend on p_0, so that $S_{p_0, \boldsymbol{y}_0} = S_{\boldsymbol{y}_0}$. In effect, this means that the characterization of the optimal s_0 in (57) holds for all solutions that share the same measurements $\boldsymbol{y}_0$. This is true, in particular, when E is strictly convex, by a standard argument in convex analysis. The description of $s_0 \in \mathcal{U}'$ as the $(\mathcal{U}', \mathcal{U})$-conjugate of a common $\widetilde{\nu}_0 \in \mathcal{N}_{\widetilde{\boldsymbol{\nu}}}$, as well as (43), (44), and (45), then follow from Theorem 1.

For the special scenario $\mathcal{N}_{\boldsymbol{p}^*} \subset \mathcal{N}_{\boldsymbol{\nu}}$, we select $\widetilde{\mathbf{V}}$ such that $\boldsymbol{p}^* = \widetilde{\mathbf{V}}^T \boldsymbol{\nu}$ (see the second part of Proposition 2) and are then able to obtain the expansion of coefficients of $p_0 = \mathrm{Proj}_{\mathcal{N}_{\boldsymbol{p}}} \{f_0\}$ directly from the measurements as $\boldsymbol{b} = \boldsymbol{p}^*(f_0) = \widetilde{\mathbf{V}}^T \boldsymbol{y}_0$, which establishes this part of the solution as well for all $f_0 \in S$. $\quad \square$

When the (semi)-norm $\| \cdot \|_{\mathcal{U}'}$ is strictly convex, Theorem 3 states that the unique solution f_0 lives in a finite-dimensional manifold that is parameterized by $\boldsymbol{b} \in \mathbb{R}^{N_0}$ (for the null-space component p_0) and $\boldsymbol{a} \in \mathbb{R}^{M-N_0}$ (for the preimage $\widetilde{\nu}_0$ of the penalized component $s_0 = (\widetilde{\nu}_0)^*$). While the two primary expansions are linear, the high-level ingredient of the representation is the duality mapping, which introduces a nonlinearity in the non-Hilbert scenario. At any rate, the main point is that the intrinsic dimensionality of the solution space is still M, as in the case of Theorem 1, except that the repartition is now very different, with the contribution of the null-space component $p_0 \in \mathcal{N}_{\boldsymbol{p}}$ being maximized since it is no longer penalized.

An important outcome of Theorem 3 is that it becomes possible to characterize the full solution set S via the specification of a single pair $(p_0, \widetilde{\nu}_0) \in \mathcal{N}_{\boldsymbol{p}} \times \mathcal{U}$. For the challenging cases where there are multiple solutions (third scenario), this requires the additional assumption that $\mathcal{N}_{\boldsymbol{p}^*} \subset \mathcal{N}_{\boldsymbol{\nu}}$, which has the desirable effect of decoupling the determination p_0 from that of $\widetilde{\nu}_0$. It turns out that this decoupling is applicable to most practical problems that involve a semi-norm regularization. The key is that it is generally possible to adapt the semi-norm topology to the problem at hand by selecting a biorthonormal system $(\boldsymbol{p}^*, \boldsymbol{p})$ with $p_1^*, \ldots, p_{N_0}^* \in \mathrm{span}\{\nu_m\}_{m=1}^M \subset \mathcal{X}$ (see, for instance, [20,58]).

5.2. Connection with prior works

The result for Hilbert spaces in Theorem 3 is well known, although it is rarely described in this form where the parameterization is tight. The result for Banach spaces is new, to the best of our knowledge. The result for the non-strictly convex case overlaps that of two recent papers by Bredies-Carioni [19] and Boyer et al. [23], in which these authors independently establish the existence of solutions of the form (45) in similar scenarios. The novel element here is the statement about the form of all extremal points as well as the identification of the configurations where p_0 is unique. Finally, we are not aware of any prior work (except [42]) where these various scenarios have been unified.

For the particular scenario where $\mathrm{T} = \mathrm{L}^{-1}$ is an isomorphism from $\mathcal{M}(\mathbb{R}^d)$ (the space of Radon measures on $\mathbb{R}^d$) to $\mathcal{U}' = \mathrm{T}\big(\mathcal{M}(\mathbb{R}^d)\big)$ and $\mathcal{X}' = \mathcal{U}' \oplus \mathcal{N}_{\boldsymbol{p}}$ with $\|f\|_{\mathcal{U}'} = \|\mathrm{L}f\|_{\mathcal{M}}$, we recover the results of [20] (when the operator L is spline-admissible) and [18, Theorem 1] by observing that the generic form of the extremal points of the unit regularization ball are $e_k = \mathrm{T}\{\delta(\cdot - \boldsymbol{x}_k)\}$ with $\boldsymbol{x}_k \in \mathbb{R}^d$. The key there is that the extremal points of the unit ball in $\mathcal{M}(\mathbb{R}^d)$ are the signed shifted Dirac measures $\{\pm\delta(\cdot - \boldsymbol{x})\}_{\boldsymbol{x} \in \mathbb{R}}$, which are then isometrically mapped into $\mathcal{U}'$ in accordance with Proposition 1. To make this more concrete, we recall that an operator $\mathrm{L} : \mathcal{X}' \to \mathcal{M}(\mathbb{R}^d)$ is called spline-admissible if

1. it is linear shift-invariant;
2. it has a finite-dimensional null space $\mathcal{N}_{\mathrm{L}} = \{f \in \mathcal{X}' : \mathrm{L}f = 0\} = \mathrm{span}\{p_n\}_{n=1}^{N_0}$;
3. it admits a Green's function $\rho_{\mathrm{L}} : \mathbb{R}^d \to \mathbb{R}$ (of slow growth) such that $\rho_{\mathrm{L}} = \mathrm{L}^{-1}\{\delta\}$.

In that scenario, we find[1] that the extremal points of (41) with $\|f\|_{\mathcal{U}'} = \|\mathrm{L}f\|_{\mathcal{M}}$ can be represented as

$$f_0 : \mathbb{R}^d \to \mathbb{R} : \boldsymbol{x} \mapsto \sum_{n=1}^{N_0} \tilde{b}_n p_n(\boldsymbol{x}) + \sum_{k=1}^{K_0} a_k \rho_{\mathrm{L}}(\boldsymbol{x} - \boldsymbol{x}_k) \tag{58}$$

with $K_0 \leq (M - N_0)$, which is the generic form of a non-uniform L-spline with knots at the $\boldsymbol{x}_k \in \mathbb{R}^d$ [59,60][61, Chapter 6]. For instance, for $\mathrm{L} = \mathrm{D}$ (the derivative operator with $d = 1$), $N_0 = 1$ with $p_1 = 1$ (the constant function), while $\rho_{\mathrm{D}}(x)$ is the unit step (Heaviside function). It follows that f_0 given by (58) is piecewise-constant with jumps at the $x_k \in \mathbb{R}$.

The setting of Theorem 3 can also be extended to the multicomponent scenarios investigated in Section 4. In essence, one can replace regularization norms by semi-norms, which then adds some corresponding null-space component(s) to the generic form of the solution. Hybrid splines, which can be seen as the continuous-domain counterpart of the multi-dictionary approach of Section 4.2, provide a powerful example of such composite modeling [62]. Another useful option is the spline variant of Corollary 2, which adds a smooth component (with a corresponding L_2 regularization) to the solution specified by (58). However, the fitting of

---

[1] There is one subtle point in the derivation (see [20]) because the correct inverse operator $\mathrm{T} = (\mathrm{Id} - \mathrm{Proj}_{\mathcal{N}_p})\mathrm{L}^{-1} = \mathrm{L}_{\boldsymbol{p}^*}^{-1}$ must incorporate a projection onto $\mathcal{U}'$. This projection depends upon the topology. This yields extremal points of the form $e_k = (\rho_{\mathrm{L}}(\cdot - \boldsymbol{x}_k) - q_k)$ with $q_k = \mathrm{Proj}_{\mathcal{N}_p}\{\rho_{\mathrm{L}}(\cdot - \boldsymbol{x}_k)\} \in \mathcal{N}_{\boldsymbol{p}}$, which translates into $\tilde{b}_n$ in (58) being different from b_n in (45).

such augmented models to data is trickier: when the intersection of the null spaces is nontrivial, it requires the specification of additional boundary conditions to ensure that the decomposition is unique [62,63].

Acknowledgments

Appendix A. Proof of the last statement in Theorem 1

We already mentioned that this characterization can be derived from Theorem 3.1 of Boyer et al. [23] by viewing an extreme point of the solution set as the degenerate case of a face with dimension $j = 0$. The proof presented in [23] relies on an earlier theorem by Klee [64], which is itself based on a foundational result by Dubins on the extreme points of the intersection of a convex set and a series of hyperplanes [65]. Here, we have chosen the latter as our starting point in order to simplify the argumentation.

Theorem 4 *(Main result of [65]). Consider a topological vector space V over the field of real numbers, a closed and bounded convex set $C \subset V$ and M hyperplanes $H_1, \ldots, H_M \subset V$. Then, any extreme point of $C \cap \left(\bigcap_{m=1}^{M} H_m \right)$ can be written as a convex combination of at most $M + 1$ extreme points of C.*

For the Banach space $\mathcal{X}'$, we denote the unit ball of size β as $B_{\mathcal{X}',\beta} = \{ f \in \mathcal{X}' : \|f\|_{\mathcal{X}'} \le \beta \}$. It then directly follows from Theorem 4 that any extreme point of

$$\mathcal{S} = B_{\mathcal{X}',\beta} \cap \{ f \in \mathcal{X}' : \boldsymbol{\nu}(f) = \boldsymbol{y} \}$$

can be written as a convex combination of at most $M + 1$ extreme points of $B_{\mathcal{X}',\beta}$. In what follows, we show that, if

$$\beta = \min_{f \in \mathcal{X}'} \|f\|_{\mathcal{X}'} \quad \text{s.t.} \quad \boldsymbol{\nu}(f) = \boldsymbol{y},$$

then any extreme points f_0 of $\mathcal{S}$ has the expansion

$$f_0 = \sum_{k=1}^{K} c_k f_k, \quad K \le M, \tag{59}$$

where $f_k \in \text{Ext}(B_{\mathcal{X}',\beta})$, and $c_k > 0$ with $\sum_{k=1}^{K} c_k = 1$. The connection with (4) is that (59) is obviously also expressible in terms of the basis vectors $e_k = f_k / \beta$ which, due to the homogeneity property of the norm, are extremal points of the unit ball in $\mathcal{X}'$.

Assume by contradiction that $K = M + 1$ and that the set $\{f_1, \ldots, f_{M+1}\}$ is linearly independent. The set of vectors $\{\boldsymbol{\nu}(f_1), \ldots, \boldsymbol{\nu}(f_{M+1})\} \subseteq \mathbb{R}^M$ is clearly linearly dependent. Hence, there exists $(\alpha_m)_{m=1}^{M+1} \ne \boldsymbol{0}$ such that

$$\boldsymbol{\nu}\left(\sum_{m=1}^{M+1} \alpha_m \boldsymbol{f}_m \right) = \sum_{m=1}^{M+1} \alpha_m \boldsymbol{\nu}(f_m) = \boldsymbol{0}. \tag{60}$$

Denote $A = \sum_{m=1}^{M+1} \alpha_m$ and consider the function $f_\epsilon = f_0 + \epsilon \sum_{m=1}^{M+1} \alpha_m f_m$ for $\epsilon \in \mathbb{R}$. On one hand, for all values of ϵ with $|\epsilon| < \epsilon_{\max} = \frac{\min_m c_m}{\max_m |\alpha_m|}$, the function

$$\frac{f_\epsilon}{1 + \epsilon A} = \sum_{m=1}^{M+1} \frac{c_m + \epsilon \alpha_m}{1 + \epsilon A} f_m$$

is in the convex hull of $\{f_1, \ldots, f_{M+1}\}$. Consequently, $\|f_\epsilon\|_{\mathcal{X}'} \leq |1 + \epsilon|\beta$. On the other hand, due to (60), we have

$$\boldsymbol{\nu}(f_\epsilon) = \boldsymbol{\nu}(f_0) + \epsilon \sum_{m=1}^{M+1} \alpha_m \boldsymbol{y}_m = \boldsymbol{y}.$$

Hence, due to the optimality of f_0, we deduce that

$$|1 + \epsilon A|\beta \geq \|f_\epsilon\|_{\mathcal{X}'} \geq \|f_0\|_{\mathcal{X}'} = \beta, \quad \forall \epsilon \in (-\epsilon_{\max}, \epsilon_{\max}).$$

This yields that $|1 + \epsilon A| \geq 1$ for all $\epsilon \in (-\epsilon_{\max}, \epsilon_{\max})$, which implies that $A = 0$. Consequently, $f_\epsilon \in \mathcal{S}$ for all $\epsilon \in (-\epsilon_{\max}, \epsilon_{\max})$. Now, since f_0 is an extreme point of $\mathcal{S}$, we deduce from $f_0 = \frac{f_{-\epsilon} + f_\epsilon}{2}$ that $f_0 = f_\epsilon$ for all $\epsilon \in (-\epsilon_{\max}, \epsilon_{\max})$, and hence, $\sum_{m=1}^{M+1} \alpha_m f_m = 0$. This is in contradiction with the linear independence of $\{f_1, \ldots, f_{M+1}\}$.

References

[1] B. Schölkopf, A.J. Smola, Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond, MIT Press, 2002.
[2] A. Berlinet, C. Thomas-Agnan, Reproducing Kernel Hilbert Spaces in Probability and Statistics, Vol. 3, Kluwer Academic, Boston, 2004.
[3] G. Kimeldorf, G. Wahba, A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, Ann. Math. Stat. 41 (2) (1970) 495–502.
[4] A. Badoual, J. Fageot, M. Unser, Periodic splines and Gaussian processes for the resolution of linear inverse problems, IEEE Trans. Signal Process. 66 (22) (2018) 6047–6061.
[5] E. De Vito, L. Rosasco, A. Caponnetto, M. Piana, A. Verri, Some properties of regularized kernel methods, J. Mach. Learn. Res. 5 (2004) 1363–1390.
[6] H. Wendland, Scattered Data Approximations, Cambridge University Press, 2005.
[7] T. Hofmann, B. Schölkopf, A.J. Smola, Kernel methods in machine learning, Ann. Stat. 36 (3) (2008) 1171–1220.
[8] B. Schölkopf, R. Herbrich, A.J. Smola, A generalized representer theorem, in: D. Helmbold, B. Williamson (Eds.), Computational Learning Theory, Springer, Berlin, Heidelberg, 2001, pp. 416–426.
[9] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.
[10] E.J. Candès, J. Romberg, Sparsity and incoherence in compressive sampling, Inverse Probl. 23 (3) (2007) 969–985.
[11] A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Rev. 51 (1) (2009) 34–81.
[12] M. Elad, Sparse and Redundant Representations. From Theory to Applications in Signal and Image Processing, Springer, 2010.
[13] Y.C. Eldar, G. Kutyniok, Compressed Sensing: Theory and Applications, Cambridge University Press, Cambridge, 2012.
[14] S. Foucart, H. Rauhut, A Mathematical Introduction to Compressive Sensing, Springer, 2013.
[15] M. Unser, J. Fageot, H. Gupta, Representer theorems for sparsity-promoting ℓ_1 regularization, IEEE Trans. Inf. Theory 62 (9) (2016) 5167–5180.
[16] E.J. Candès, C. Fernandez-Granda, Towards a mathematical theory of super-resolution, Commun. Pure Appl. Math. 67 (6) (2014) 906–956.
[17] V. Duval, G. Peyré, Exact support recovery for sparse spikes deconvolution, Found. Comput. Math. 15 (5) (2015) 1315–1355.
[18] A. Flinth, P. Weiss, Exact solutions of infinite dimensional total-variation regularized problems, Inf. Inference (2018) 1–37.
[19] K. Bredies, M. Carioni, Sparsity of solutions for variational inverse problems with finite-dimensional data, Calc. Var. Partial Differ. Equ. 59 (14) (2020) 26.
[20] M. Unser, J. Fageot, J.P. Ward, Splines are universal solutions of linear inverse problems with generalized-TV regularization, SIAM Rev. 59 (4) (2017) 769–793.
[21] H. Gupta, J. Fageot, M. Unser, Continuous-domain solutions of linear inverse problems with Tikhonov versus generalized TV regularization, IEEE Trans. Signal Process. 66 (17) (2018) 4670–4684.
[22] M. Unser, A unifying representer theorem for inverse problems and machine learning, Found. Comput. Math. (2020), https://doi.org/10.1007/s10208-020-09472-x.
[23] C. Boyer, A. Chambolle, Y. De Castro, V. Duval, F. De Gournay, P. Weiss, On representer theorems and convex regularization, SIAM J. Optim. 29 (2) (2019) 1260–1281.
[24] N. Aronszajn, Theory of reproducing kernels, Trans. Am. Math. Soc. 68 (3) (1950) 337–404.
[25] R.E. Megginson, An Introduction to Banach Space Theory, Springer, New York, 1998.
[26] C.A. Micchelli, M. Pontil, On learning vector-valued functions, Neural Comput. 17 (1) (2005) 177–204.
[27] M.A. Alvarez, L. Rosasco, N.D. Lawrence, Kernels for vector-valued functions: a review, Found. Trends Mach. Learn. 4 (3) (2012) 195–266.

[28] M. Gönen, E. Alpaydın, Multiple kernel learning algorithms, J. Mach. Learn. Res. 12 (Jul) (2011) 2211–2268.
[29] C.A. Micchelli, M. Pontil, Learning the kernel function via regularization, J. Mach. Learn. Res. 6 (Jul) (2005) 1099–1125.
[30] M.A. van Wyk, T.S. Durrani, A framework for multiscale and hybrid RKHS-based approximators, IEEE Trans. Signal Process. 48 (12) (2000) 3559–3568.
[31] C. de Boor, R.E. Lynch, On splines and their minimum properties, J. Math. Mech. 15 (6) (1966) 953–969.
[32] J. Duchon, Splines minimizing rotation-invariant semi-norms in Sobolev spaces, in: W. Schempp, K. Zeller (Eds.), Constructive Theory of Functions of Several Variables, Springer-Verlag, Berlin, 1977, pp. 85–100.
[33] A.Y. Bezhaev, V.A. Vasilenko, Variational Theory of Splines, Kluwer Academic/Plenum Publishers, New York, 2001.
[34] A. Mosamam, J. Kent, Semi-reproducing kernel Hilbert spaces, splines and increment kriging, J. Nonparametr. Stat. 22 (6) (2010) 711–722.
[35] M. Elad, A. Bruckstein, A generalized uncertainty principle and sparse representation in pairs of bases, IEEE Trans. Inf. Theory 48 (9) (2002) 2558–2567.
[36] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Inf. Theory 49 (12) (2003) 3320–3325.
[37] J.-L. Starck, M. Elad, D. Donoho, Redundant multiscale transforms and their application for morphological component separation, in: Advances in Imaging and Electron Physics, Elsevier, 2004, pp. 287–348.
[38] J.-J. Fuchs, On sparse representations in arbitrary redundant bases, IEEE Trans. Inf. Theory 50 (6) (2004) 1341–1344.
[39] H. Rauhut, K. Schnass, P. Vandergheynst, Compressed sensing and redundant dictionaries, IEEE Trans. Inf. Theory 54 (5) (2008) 2210–2219.
[40] E. Candès, Y. Eldar, D. Needell, P. Randall, Compressed sensing with coherent and redundant dictionaries, Appl. Comput. Harmon. Anal. 31 (1) (2011) 59–73.
[41] J. Lin, S. Li, Y. Shen, Compressed data separation with redundant dictionaries, IEEE Trans. Inf. Theory 59 (7) (2013) 4309–4315.
[42] M. Unser, A representer theorem for deep neural networks, J. Mach. Learn. Res. 20 (110) (2019) 1–30.
[43] G. Wahba, Spline Models for Observational Data, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1990.
[44] W. Rudin, Functional Analysis, 2nd edition, McGraw–Hill Series in Higher Mathematics, McGraw–Hill, New York, 1991.
[45] A. Beurling, A.E. Livingston, A theorem on duality mappings in Banach spaces, Ark. Mat. 4 (5) (1962) 405–411.
[46] I. Cioranescu, Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems, Vol. 62, Springer Science & Business Media, 2012.
[47] T. Schuster, B. Kaltenbacher, B. Hofmann, K.S. Kazimierski, Regularization Methods in Banach Spaces, Vol. 10, Walter de Gruyter, 2012.
[48] M. Riesz, Sur les fonctions conjuguées, Math. Z. 27 (1927) 218–244.
[49] W. Rudin, Functional Analysis, McGraw–Hill Series in Higher Mathematics, McGraw–Hill, New York, 1973.
[50] V. Chandrasekaran, B. Recht, P.A. Parrilo, A.S. Willsky, The convex geometry of linear inverse problems, Found. Comput. Math. 12 (6) (2012) 805–849.
[51] F.L. Bauer, J. Stoer, C. Witzgall, Absolute and monotonic norms, Numer. Math. 3 (1) (1961) 257–264.
[52] P.N. Dowling, S. Saejung, Extremal structure of the unit ball of direct sums of Banach spaces, Nonlinear Anal., Theory Methods Appl. 68 (4) (2008) 951–955.
[53] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006.
[54] H. Zhang, Y. Xu, J. Zhang, Reproducing kernel Banach spaces for machine learning, J. Mach. Learn. Res. 10 (2009) 2741–2775.
[55] H. Zhang, J. Zhang, Regularized learning in Banach spaces as an optimization problem: representer theorems, J. Glob. Optim. 54 (2) (2012) 235–250.
[56] E.J. Candès, C. Fernandez-Granda, Super-resolution from noisy data, J. Fourier Anal. Appl. 19 (6) (2013) 1229–1254.
[57] Q. Denoyelle, V. Duval, G. Peyré, Support recovery for sparse super-resolution of positive measures, J. Fourier Anal. Appl. 23 (5) (2017) 1153–1194.
[58] M. Unser, J. Fageot, Native Banach spaces for splines and the formulation of variational inverse problems, arXiv Preprint, URL arXiv:1904.10818.
[59] M.H. Schultz, R.S. Varga L-splines, Numer. Math. 10 (4) (1967) 345–369.
[60] L.L. Schumaker, Spline Functions: Basic Theory, 3rd edition, Cambridge Univ. Press, Cambridge, 2007.
[61] M. Unser, P.D. Tafti, An Introduction to Sparse Stochastic Processes, Cambridge University Press, 2014.
[62] T. Debarre, S. Aziznejad, M. Unser, Hybrid-spline dictionaries for continuous-domain inverse problems, IEEE Trans. Signal Process. 67 (22) (2019) 5824–5836.
[63] T. Debarre, S. Aziznejad, M. Unser, Continuous-domain formulation of inverse problems for composite sparse-plus-smooth signals, arXiv preprint, arXiv:2103.13380.
[64] V. Klee, On a theorem of Dubins, J. Math. Anal. Appl. 7 (3) (1963) 425–427.
[65] L.E. Dubins, On extreme points of convex sets, J. Math. Anal. Appl. 5 (2) (1962) 237–244.