

# On Representer Theorems and Convex Regularization

Claire Boyer, Antonin Chambolle, Yohann de Castro, Vincent Duval, Frédéric de Gournay, Pierre Weiss

▶ **To cite this version:**

Claire Boyer, Antonin Chambolle, Yohann de Castro, Vincent Duval, Frédéric de Gournay, et al.. On Representer Theorems and Convex Regularization. SIAM Journal on Optimization, 2019, 29 (2), pp.1260-1281. 10.1137/18M1200750 . hal-01823135v3

**HAL Id: hal-01823135**
**https://hal.science/hal-01823135v3**

Submitted on 26 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# On Representer Theorems and Convex Regularization

Claire Boyer[1], Antonin Chambolle[2], Yohann De Castro[3,4], Vincent Duval[4,5], Frédéric de Gournay[6] and Pierre Weiss[7].

[1] LPSM, Sorbonne Université, ENS Paris [2] CMAP, CNRS, Polytechnique, Palaiseau
[3] LMO, Université Paris-Sud, Orsay [4] MOKAPLAN, INRIA, Paris
[5] Ceremade, Université Paris Dauphine [6] INSA, Université de Toulouse
[7] ITAV, CNRS, Université de Toulouse

November 26, 2018

**Abstract**

We establish a general principle which states that regularizing an inverse problem with a convex function yields solutions which are convex combinations of a small number of *atoms*. These atoms are identified with the extreme points and elements of the extreme rays of the regularizer level sets. An extension to a broader class of quasi-convex regularizers is also discussed. As a side result, we characterize the minimizers of the total gradient variation, which was still an unresolved problem.

**Keywords:** Inverse problems, Convex regularization, Representer theorem, Vector space, Total variation

## 1 Introduction

Let $E$ denote a real vector space. Let $\Phi : E \to \mathbb{R}^m$ be a linear mapping called *sensing operator* and $u \in E$ denote a signal. The main results in this paper describe the structural properties of certain solutions of the following problem:

$$\inf_{u \in E} f(\Phi u) + R(u), \tag{1}$$

where $R : E \to \mathbb{R} \cup \{+\infty\}$ is a convex function called *regularizer* and $f$ is an arbitrary convex or non-convex function called *data fitting term*. In many applications, one looks for "sparse solutions" that are linear sums of a few atoms. This article investigates the theoretical legitimacy of this usage.

**Representer theorems and Tikhonov regularization** The name *representer theorem* comes from the field of machine learning [43]. To provide a first concrete example[1], assume that $\Phi \in \mathbb{R}^{m \times n}$ is a finite dimensional measurement operator and $L \in \mathbb{R}^{p \times n}$ is a linear transform. Solving an inverse problem using Tikhonov regularization amounts to finding the minimizers of

$$\min_{u \in \mathbb{R}^m} \frac{1}{2}\|\Phi u - y\|_2^2 + \frac{1}{2}\|Lu\|_2^2. \tag{2}$$

Provided that $\ker\Phi \cap \ker L = \{0\}$, it is possible to show that, whatever the data $y$ is, solutions are always of the form

$$u^\star = \sum_{i=1}^{m} \alpha_i \psi_i + u_K, \tag{3}$$

where $u_K \in \ker(L)$ and $\psi_i = (\Phi^T\Phi + L^TL)^{-1}(\phi_i)$, where $\phi_i^T \in \mathbb{R}^n$ is the $i$-th row of $\Phi$. This result characterizes structural properties of the minimizers without actually needing to solve the problem. In addition, when $E$ is an infinite dimensional Hilbert space, Equation (3) sometimes allows to compute exact solutions, by simply solving a finite dimensional linear system. This is a critical observation that explains the practical success of kernel methods and radial basis functions [49].

**Representer theorems and convex regularization** The Tikhonov regularization (2) is a powerful tool when the number $m$ of observations is large and the operator $\Phi$ is not too ill-conditioned. However, recent results in the fields of compressed sensing [17], matrix completion [11] or super-resolution [46, 10] - to name a few - suggest that much better results may be obtained in general, by using convex regularizers, with level sets containing singularities. Popular examples of regularizers in the finite dimensional setting include the indicator of the nonnegative orthant [18], the $\ell^1$-norm [17] or its composition with a linear operator [41] and the nuclear norm [11]. Those results were nicely unified in [13] and one of the critical arguments behind all these techniques is a representer theorem of type (3). In most situations however, this argument is only implicit. The

[1]Here, we follow the presentation of [27].

main objective of this paper is to state a generalization of (3) to arbitrary convex functions $R$. It covers all the aforementioned problems, but also new ones for problems formulated over the space of measures.

To the best of our knowledge, the name "*representer theorem*" is new in the field of convex regularization and its first mention is due to Unser, Fageot and Ward in [48]. Describing the solutions of (1) is however an old problem which has been studied since at least the 1940's in the case of Radon measure recovery.

**Total variation regularization of Radon measures** A typical example of inverse problem in the space of measures is

$$\min_{\mu \in \mathcal{M}(\Omega)} |\mu|(\Omega) \quad \text{s.t.} \quad \Phi\mu = y \tag{4}$$

where $\Omega \subseteq \mathbb{R}^N$, $\mathcal{M}(\Omega)$ denotes the space of Radon measures, $|\mu|(\Omega)$ is the total variation of the measure $\mu$ (see Section 4) and $\Phi\mu$ is a vector of *generalized moments*, *i.e.* $\Phi\mu = \left(\int_\Omega \varphi_i(x)\mathrm{d}\mu(x)\right)_{1\leq i\leq m}$ where $\{\varphi_i\}_{1\leq i\leq m}$ is a family of continuous functions (which "vanish at infinity" if $\Omega$ is not compact).

Problems of the form (4) have received considerable attention since the pioneering works of Beurling [6] and Krein [31], sometimes under the name *L-moment problem* (see the monograph [32]). To the best of our knowledge, the first "representer theorem" for problems of the form (1) is given for (4) by Zuhovickiĭ [50] (see [51, Th. 3] for an English version). It essentially states that

$$\textit{There exists a solution to (4) of the form } \sum_{i=1}^{r} a_i \delta_{x_i}, \textit{ with } r \leq m. \tag{5}$$

A more precise result was given by Fisher and Jerome in [23]. When considering the problem (4), and for a bounded domain $\Omega$, the result reads as follows:

$$\textit{The extreme points of the solution set to (4) are of the form} \quad \sum_{i=1}^{r} a_i \delta_{x_i}, \textit{ with } r \leq m. \tag{6}$$

Incidentally, the Fisher-Jerome theorem considers more general problems of the form:

$$\min_{u \in E} |Lu|(\Omega) \quad \text{s.t.} \quad Lu \in \mathcal{M}(\Omega) \quad \text{and} \quad \Phi u = y, \tag{7}$$

where $E \subseteq \mathcal{D}'(\Omega)$ is a suitably defined Banach space of distributions, $L : \mathcal{D}'(\Omega) \to \mathcal{D}'(\Omega)$ maps $E$ onto $\mathcal{M}(\Omega)$ and $\Phi : E \to \mathbb{R}^m$ is a continuous linear operator. We refer to Section 4 for precise assumptions. Let us mention that the initial results by Fisher-Jerome were extended to a significantly more general setting in [48].

It is important to note that the Fisher-Jerome theorem [23] provides a much finer description of the solution set than Zuhovickiĭ's result [50]. Indeed, the well-known Krein-Milman theorem states that, if $E$ is endowed with the topology of a locally convex Hausdorff vector space and $C \subset E$ is compact convex, then $C$ is the closed convex hull of its extreme points,

$$\operatorname{cl}\operatorname{conv}(\operatorname{ext}(C)) = C. \tag{8}$$

In other words, the solutions described by the Fisher-Jerome theorem are sufficient to recover *the whole set of solutions*. Let us mention that the Krein-Milman theorem was extended by Klee [30] to unbounded sets: if $C$ is locally compact, closed, convex, and contains no line, then

$$\operatorname{cl}\operatorname{conv}(\operatorname{ext}(C) \cup \operatorname{rext}(C)) = C, \tag{9}$$

where $\operatorname{rext}(C)$ denotes the union of the extreme rays of $C$ (see Section 2 below).

**"Representer theorems" for convex sets** As the Dirac masses are the extreme points of the total variation unit ball, each of the above-mentioned "representer theorems" for inverse problems actually reflect some phenomenon in the geometry of convex sets. In that regard, the celebrated Minkowski-Carathéodory theorem [28, Th. III.2.3.4] is fundamental: any point of a compact convex set in an $m$-dimensional space is a convex combination of (at most) $m + 1$ of its extreme points. In [29, Th. (3)], Klee removed the boundedness assumption and obtained the following extension: any point of a closed convex set in an $m$-dimensional space is a convex combination of (at most) $m + 1$ extreme points, or $m$ points, each an extreme point or a point in an extreme ray.

One purpose of the present paper is to point out the connection between the Fisher-Jerome theorem and a lesser known theorem by Dubins [20] (see also [8, Exercise II.7.3.f]):

*The extreme points of the intersection of $C$ with an affine space of codimension $m$ are convex combination of (at most)*[2] *$m + 1$ extreme points of $C$,*

[2] In the rest of the paper, we omit the mention "at most", with the convention that some points may be chosen identical.

provided $C$ is linearly bounded and linearly closed (see Section 2). That theorem was extended by Klee [29] to deal with the unbounded case.

Although the connection with the Fisher-Jerome is striking, Dubins' theorem actually provides one extreme point too many. In the case of (4), it would yield two Dirac masses for one linear measurement. We provide in this paper a refined analysis of the case of variational problems, which ensures at most $m$ extreme points.

**Contributions** The main results of this paper yield a description of some solutions to (1) of the following form:

$$u^\star = \sum_{i=1}^{r} \alpha_i \psi_i + u_K,$$

where $r \leq m$, the atoms $\psi_i$ are identified with some extreme points (or points in extreme rays) of the regularizer level sets, and $u_K$ is an element of the so-called constancy space of $R$, *i.e.* the set of directions along which $R$ is invariant. The results take the form (5), when $f$ is an arbitrary function and the form (6) when it is convex. We provide tight bounds on the number of atoms $r$ that depend on the geometry of the level sets and on the link between the constancy space of $R$ and the measurement operator $\Phi$.

Our general theorems then allow us to revisit many results of the literature (linear programming, semi-definite programming, nonnegative constraints, nuclear norm, analysis priors), yielding simple and accurate descriptions of the minimizers. Our analysis also allows us to characterize the solutions of a resisting problem: we provide a representation theorem for the minimizers of the total gradient variation [41] as sums of indicators of simple sets. This provides a simple explanation to the staircaising effect when only a few measurements are used.

Let us mention that, shortly after this work was posted on arXiv, similar results appeared, with somewhat different proofs, in a paper by Bredies and Carioni [9].

## 2 Notation and Preliminaries

Throughout the paper, unless otherwise specified, $E$ denotes a finite or infinite dimensional real vector space and $C \subseteq E$ is a convex set. Given two distinct points $x$ and $y$ in $E$, we let $]x, y[ = \{tx + (1-t)y \ : \ 0 < t < 1\}$ and $[x, y] = \{tx + (1-t)y : 0 \leq t \leq 1\}$ denote the open and closed segments

joining $x$ to $y$. We recall the following definitions, and we refer to [20, 30] for more details.

**Lines, rays, and linearly closed sets** A *line* is an affine subspace of $E$ with dimension 1. An open half-line, *i.e.* a set of the form $\rho = \{p + tv : t > 0\}$, where $p, v \in E$, $v \neq 0$, is called a *ray* (through $p$). We say that the set $C$ is *linearly closed* (resp. linearly bounded) if the intersection of $C$ and a line of $E$ is closed (resp. bounded) for the natural topology of the line. If $E$ is a topological vector space and $C$ is closed for the corresponding topology, then $C$ is linearly closed.

If $C$ is linearly closed and contains some ray $\rho = p+\mathbb{R}_+^* v$, it also contains the endpoint $p$ as well as the rays $q + \mathbb{R}_+ v$ for all $q \in C$. Therefore, if $C$ contains a ray (resp. line), it recesses in the corresponding direction.

**Recession cone and lineality space** The set of all $v \in E$ such that $C + \mathbb{R}_+^* v \subseteq C$ is a convex cone called the *recession cone of* $C$, which we denote by $\mathrm{rec}(C)$. If $C$ is linearly closed then so is $\mathrm{rec}(C)$, and $\mathrm{rec}(C)$ is the union of $0$ and all the vectors $v$ which direct the rays of $C$. In particular, $C$ contains a line if and only the vector space

$$\mathrm{lin}(C) \stackrel{\text{def.}}{=} \mathrm{rec}(C) \cap (-\mathrm{rec}(C)) \tag{10}$$

is non trivial. The vector space $\mathrm{lin}(C)$ is called the *lineality space* of $C$. It corresponds to the largest vector space of invariant directions for $C$.

If $E$ is finite dimensional and $C$ is closed, the recession cone coincides with the *asymptotic cone*.

**Extreme points, extremal rays, faces** An *extreme point* of $C$ is a point $p \in C$ such that $C \setminus \{p\}$ is convex. An *extremal ray* of $C$ is a ray $\rho \in C$ such that if $x, y \in C$ and $]x, y[$ intersects $\rho$, then $]x, y[ \subset \rho$. If $C$ contains the endpoint $p$ of $\rho$ (*e.g.* if $C$ is linearly closed), this is equivalent to $p$ being an extreme point of $C$ and $C \setminus \rho$ being convex.

Following [20, 29], if $p \in C$, the smallest face of $C$ which contains $p$ is the union of $\{p\}$ and all the open segments in $C$ which have $p$ as an inner point. We denote it by $\mathcal{F}_C(p)$. The (co)dimension of $\mathcal{F}_C(p)$ is defined as the (co)dimension of its affine hull. The collection of all elementary faces, $\{\mathcal{F}_C(p)\}_{p \in C}$, is a partition of $C$. Extreme points correspond to the zero-dimensional faces of $C$, while extreme rays are (generally a strict subcollection of the) one-dimensional faces.

**Quotient by lines** As noted above, if $C$ is linearly closed, it contains a line if and only if the vector space $\mathrm{lin}(C)$ defined in (10) is nontrivial. In that case, letting $W$ denote some linear supplement to $\mathrm{lin}(C)$ , we may write

$$C = \tilde{C} + \mathrm{lin}(C), \text{ with } \tilde{C} \overset{\text{def.}}{=} C \cap W \tag{11}$$

and the corresponding decomposition is unique (*i.e.* any element of $C$ can be decomposed in a unique way as the sum of an element of $\tilde{C}$ and $\mathrm{lin}(C)$). The convex set $\tilde{C}$ (isomorphic to the projection of $C$ onto the quotient space $E/\mathrm{lin}(C)$) is then linearly closed, and the decomposition of $C$ in elementary faces is exactly given by the partition $\{\mathcal{F}_{\tilde{C}}(p) + \mathrm{lin}(C)\}_{p\in\tilde{C}}$, where $\mathcal{F}_{\tilde{C}}(p)$ is the smallest face of $p$ in $\tilde{C}$.

One may check that $\tilde{C}$ contains no line, as its recession cone $\mathrm{rec}(\tilde{C})$, the projection of $\mathrm{rec}(C)$ onto $W$ parallel to $\mathrm{lin}(C)$, is a *salient* convex cone.

# 3 Abstract representer theorems

## 3.1 Main result

Our main result describes the facial structure of the solution set to

$$\min_{u\in E} R(u) \quad \text{s.t.} \quad \Phi u = y, \tag{$\mathcal{P}$}$$

where $y \in \mathbb{R}^m$, $\Phi : E \to \mathbb{R}^m$ is a linear operator, and $m \leq \dim E$, $m < +\infty$. In the following, let $t^\star$ denote the *optimal value* of ($\mathcal{P}$), $\mathcal{S}^\star$ denote its *solution set*, and $C^\star$ denote the *corresponding level set* of $R$,

$$C^\star \overset{\text{def.}}{=} \{u \in E \ : \ R(u) \leq t^\star\}\,. \tag{12}$$

**Theorem 1.** *Let $R : E \to \mathbb{R} \cup \{+\infty\}$ be a convex function. Assume that $\inf_E R < t^\star < +\infty$, that $\mathcal{S}^\star$ is nonempty and that the convex set $C^\star$ is linearly closed and contains no line. Let $p \in \mathcal{S}^\star$ and let $j$ be the dimension of the face $\mathcal{F}_{\mathcal{S}^\star}(p)$. Then $p$ belongs to a face of $C^\star$ with dimension at most $m + j - 1$.*

*In particular, $p$ can be written as a convex combination of:*

- *$m + j$ extreme points of $C^\star$,*
- *or $m + j - 1$ points of $C^\star$, each an extreme point of $C^\star$ or in an extreme ray of $C^\star$.*

*Moreover,* $\mathrm{rec}(\mathcal{S}^\star) = \mathrm{rec}(C^\star) \cap \ker(\Phi)$ *and therefore* $\mathrm{lin}(\mathcal{S}^\star) = \mathrm{lin}(C^\star) \cap \ker(\Phi)$.

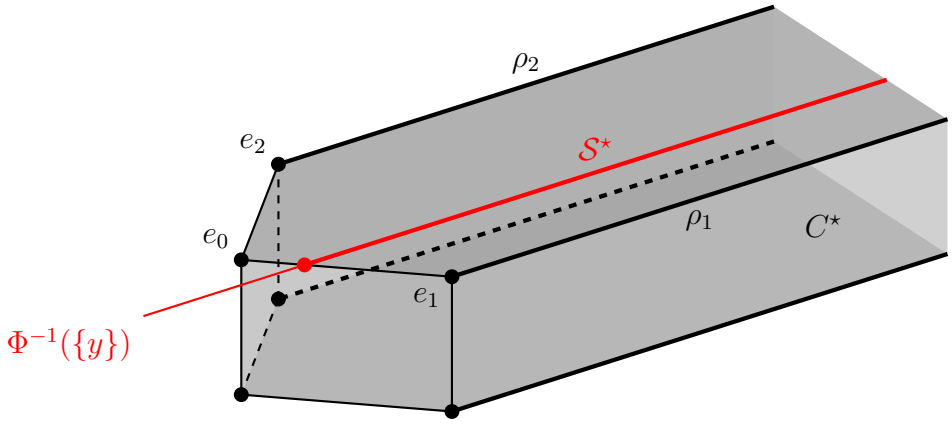



Figure 1: An illustration of theorem 1 for $m = 2$. The solution set $\mathcal{S}^\star = C^\star \cap \Phi^{-1}(\{y\})$ is made of an extreme point and an extreme ray. The extreme point is a convex combination of $\{e_0, e_1\}$. Depending on their position, the points in the ray are a convex combination of $\{e_0, e_1, e_2\}$ or a pair of points, one in $\rho_1$ and the other in $\rho_2$.

The proof of theorem 1 is given in Section 6.1. Before extending this theorem to a wider setting, let us formulate some remarks.

**Remark 1** (Extreme points and extreme rays of $\mathcal{S}^\star$)**.** *In particular* ($j = 0$), each extreme point *of $\mathcal{S}^\star$ is a convex combination of $m$ extreme points of $C^\star$, or a convex combination of $m-1$ points of $C^\star$, each an extreme point of $C^\star$ or in an extreme ray. Similarly* ($j = 1$), each point on an extreme ray *of $\mathcal{S}^\star$ is a convex combination of $m+1$ extreme points of $C^\star$, or a convex combination of $m$ points of $C^\star$, each an extreme point of $C^\star$ or in an extreme ray. Hence, provided the assumptions of Klee's theorem (see (9)) hold, Theorem 1 completely charaterizes the solution set. An illustration is provided in fig. 1.*

**Remark 2** (The hypothesis $\inf_E R < t^\star$)**.** *We have focused on the case $t^\star > \inf_E R$ in the theorem since the case $t^\star = \min R$ is easier. In that case, $M = \Phi^{-1}(\{y\})$ can be in arbitrary position (i.e. not necessarily tangent) w.r.t. $C^\star = \operatorname{argmin} R$, and one can only use the general Dubins-Klee theorem [20, 29] to describe their intersection. As a result the conclusions of theorem 1 are slightly weakened, $p$ belongs to a face of $C^\star$ with dimension $m + j$, and one must add one more point in the convex combination (e.g., for $j = 0$, each extreme point of $\mathcal{S}^\star$ is a convex combination of $m + 1$ extreme points of $C^\star$, or $m$ points...).*

**Remark 3** (Gauge functions or semi-norms)**.** *A common practice in inverse problems is to consider positively homogeneous regularizers $R$, such as (semi)-norms or gauge functions of convex sets. In that case the extreme points of $C^\star$*

*correspond, up to a rescaling, to the extreme points of* $\{u \in E : R(u) \leq 1\}$. *In several cases of interest, the extreme points of such convex sets are well understood, see Section 4 for examples in Banach spaces or, for instance, the paper [13, Sec. 2.2] for examples in finite dimensional spaces.*

**Remark 4** (Extension to semi-strictly quasi convex functions)**.** *Theorem 1 can be extended to the case where* $R$ *is a semi-strictly quasi-convex function. A function* $R$ *is said to be* semi-strictly quasi-convex *[14] if it is quasi-convex and if*

$$R(x) < R(y) \Longrightarrow R(\lambda x + (1-\lambda)y) < R(y) \quad \forall \lambda \in ]0,1[.$$

*In words, semi-strictly quasi-convex functions are functions that, when restricted to a line are successively decreasing, constant and increasing on their domain. In comparison, strictly quasi-convex functions are successively decreasing and increasing while quasi-convex functions successively non-increasing and non-decreasing.*

*The set of semi-strictly quasi-convex functions is a subset of quasi-convex functions and it contains all convex and strictly quasi-convex functions. In the proof of theorem 1, only the semi-strictly quasi-convex property is required to ensure that (28) holds.*

**Remark 5** (Topological properties)**.** *The assumption that* $C^\star$ *is linearly closed is fulfilled in most practical cases, since* $E$ *is usually endowed with the topology of a Banach (or locally convex) vector space and* $R$ *is assumed to be lower semi-continuous (so as to guarantee the existence of a solution to* $(\mathcal{P})$*). Note also that if* $R$ *is lower semi-continuous on any line (for the natural topology of the line), the set* $C^\star$ *is linearly closed.*

## 3.2 The case of level sets containing lines

The reader might be intrigued by the assumption of theorem 1 that $C^\star$ contains no line, since in several applications the regularizer $R$ is invariant by the addition of, *e.g.*, constant functions or low-degree polynomials (see Section 4). In that case, one is generally interested in the non-constant or non-polynomial part, and it is natural to consider a quotient problem for which the theorem applies. We describe below (see corollary 1) how our result extends to the case where $C^\star$ contains lines.

If $C^\star$ is linearly closed and contains some line, it is translation-invariant in the corresponding direction. The collection of all such directions is the lineality space of $C^\star$ (see Section 2), we denote it by $K \stackrel{\text{def.}}{=} \operatorname{lin}(C^\star)$ (typically, if $R$ is the composition of a linear operator and a norm, $K$ is the *kernel* of that linear operator). Let $\pi_K : E \to E/K$ be the canonical projection map.

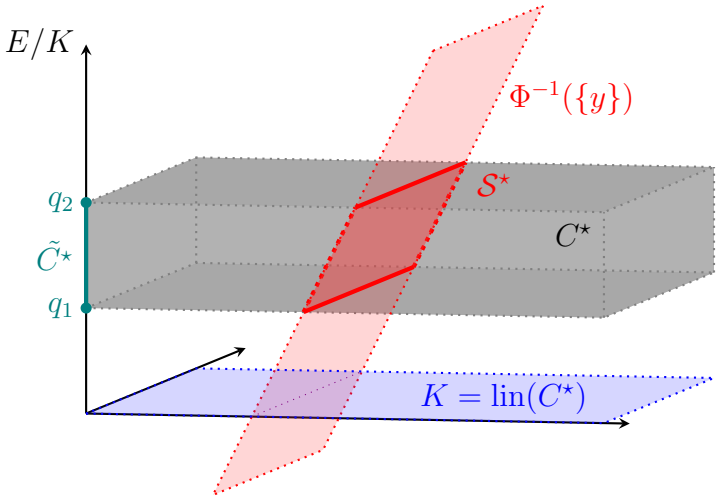



Figure 2: Taking the quotient by $K = \mathrm{lin}(C^\star)$ yields a level set $\tilde{C}^\star$ with no line. In this figure, to simplify the notation, we have omitted the isomorphism $\psi_K$ (*i.e.* in this figure $C^\star$ shoud be replaced with $\psi_K(C^\star)$, and similarly for $\mathcal{S}^\star$ and $\Phi^{-1}(\{y\})$).

We recall that there exists a linear isomorphism $\psi_K : E \to (E/K) \times K$ such that the first component of $\psi_K(p)$ is $\pi_K(p)$ for all $p \in E$. We may now describe the equivalence classes (modulo $K$) of the solutions.

**Corollary 1.** *Let $R : E \to [-\infty, +\infty]$ be a convex function. Assume that $\inf_E R < t^\star < +\infty$, that $\mathcal{S}^\star$ is nonempty and that the convex set $C^\star$ is linearly closed. Let $K \stackrel{\text{def.}}{=} \mathrm{lin}(C^\star)$ be the lineality space of $C^\star$ and $d \stackrel{\text{def.}}{=} \dim\Phi(K)$. Let $p \in \mathcal{S}^\star$, let $\pi_K(p)$ denote its equivalence class, and let $j$ be the dimension of the face $\mathcal{F}_{\pi_K(\mathcal{S}^\star)}(\pi_K(p))$.*

*Then, $\pi_K(p)$ belongs to a face of $\pi_K(C^\star)$ with dimension at most $m+j-d-1$. In particular,*

- *$\pi_K(p)$ is a convex combination of $m + j - d$ extreme points of $\pi_K(C^\star)$,*
- *or $\pi_K(p)$ is a convex combination of $m + j - d - 1$ points of $\pi_K(C^\star)$, each an extreme point of $\pi_K(C^\star)$ or in an extreme ray of $\pi_K(C^\star)$.*

*As a result, letting $\tilde{q}_1, \ldots, \tilde{q}_r$ denote those extreme points (or points in extreme rays),*

$$p = \sum_{i=1}^{r} \theta_i \psi_K^{-1}(\tilde{q}_i, 0) + u_K, \quad \textit{where} \quad \theta_i \geq 0, \ \sum_{i=1}^{r} \theta_i = 1, \quad \textit{and} \quad u_K \in K. \tag{13}$$

The proof of corollary 1 is given in Section 6.2.

One can have an explicit representation with elements of $E$ of a solution $p \in \mathcal{S}^\star$. Indeed, let $W$ be some linear complement to $K = \mathrm{lin}(C^\star)$. One may decompose $C^\star = \tilde{C}^\star + K$, where $\tilde{C}^\star = C \cap W$, and observe that $\pi_K(C^\star)$ and $\tilde{C}^\star$ are isomorphic. In this case, corollary 1 implies that $p$ can be written as the sum of one point in $\mathrm{lin}(C^\star)$ and of a convex combination of:

- $m + j - d$ extreme points of $\tilde{C}^\star$,
- or $m + j - 1 - d$ points of $\tilde{C}^\star$, each an extreme point of $\tilde{C}^\star$ or in an extreme ray of $\tilde{C}^\star$.

## 3.3 Extensions to data fitting functions

In this section, we discuss the extension of the above results to more general problems of the form

$$\inf_{u \in E} f(\Phi u) + R(u), \tag{$\mathcal{P}_f$}$$

where $f : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is an arbitrary fidelity term.

### 3.3.1 Convex data fitting term

When $f$ is a *convex* data fitting function $f$, we get the following result.

**Corollary 2.** *Assume that $f$ is convex and that the solution set $\mathcal{S}_f^\star$ of ($\mathcal{P}_f$) is nonempty. Let $p \in \mathcal{S}_f^\star$ such that $C^\star \overset{\text{def.}}{=} \{u \in E, R(u) \leq R(p)\}$ is linearly closed, $K \overset{\text{def.}}{=} \mathrm{lin} C^\star$ and let $j$ be the dimension of the face $\mathcal{F}_{\pi_K(\mathcal{S}_f^\star)}(\pi_K(p))$.*

*If $\inf_E R < R(p)$, then the conclusions of corollary 1 (or theorem 1 if $K = \{0\}$) hold.*

*If $\inf_E R = R(p)$, they hold with $1$ more dimension (see remark 2).*

Let us recall that, in view of remark 5, if $E$ is a topological vector space and $R$ is lower semi-continuous, $C^\star$ is closed regardless of the choice $p$.

*Proof.* Let $y = \Phi p$ and consider the following problem:

$$\min_{u \in E} R(u) \quad \text{s.t.} \quad \Phi u = y. \tag{$\mathcal{P}_{\{y\}}$}$$

Let $\mathcal{S}_{\{y\}}^\star$ denote its solution set. It is a convex subset of $\mathcal{S}_f^\star$, with $p \in \mathcal{S}_{\{y\}}^\star$. Additionally, if $j$ is the dimension of $\mathcal{F}_{\pi_K(\mathcal{S}_f^\star)}(p)$ (resp. $\mathcal{F}_{\mathcal{S}_f^\star}(p)$ if $C^\star$ contains

no line), then the face $\mathcal{F}_{\pi_K(\mathcal{S}^\star_{\{y\}})}(p)$ (resp. $\mathcal{F}_{\mathcal{S}^\star_{\{y\}}}(p)$) has dimension at most $j$, since $\mathcal{S}^\star_{\{y\}} \subseteq \mathcal{S}^\star_f$. It suffices to apply corollary 1 (resp. theorem 1) to obtain the result. □

**Remark 6** (The case of a strictly-convex function)**.** *In the case when $f$ is strictly convex, it is known that $\Phi\mathcal{S}^\star_f$ is a singleton, which means that $\mathcal{S}^\star_{\{y\}} = \mathcal{S}^\star_f$.*

**Remark 7** (The case of quasi-convex functions)**.** *The result actually holds whenever the solution set $\mathcal{S}^\star_f$ is convex. In particular, this property holds when $f$ is quasi-convex and $R$ is convex.*

### 3.3.2 Non-convex function

In the general case, *i.e.* when $f : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ is an arbitrary function, it is difficult to describe the structure of the solution set. However, one may choose a solution $p_0$ (as before, provided it exists) and observe that it is also a solution to ($\mathcal{P}$) for $y \overset{\text{def.}}{=} \Phi p_0$. Then, one may apply corollary 1, but the difficult part is that the dimensions $j$ to consider are with respect to the solution set $\mathcal{S}^\star$ of the *convex* problem ($\mathcal{P}$). Nevertheless, if one is able to assert that the solution set $\mathcal{S}^\star$ has at least an extreme point $p$, then corollary 1 ensures that $p$ can be written in the form (13), where $r \leq m$ and the $\tilde{q}_i$'s are extreme points (or points in extreme rays) of $C^\star$. Since $p$ must also be a solution to ($\mathcal{P}_f$), one obtains that there exists a solution to ($\mathcal{P}_f$) of the form (13).

## 3.4 Ensuring the existence of extreme points

It is important to note that, in theorem 1, the existence a face of $\mathcal{S}^\star$ with dimension $j$ is not guaranteed (nor, for $j = 0$, the existence of extreme points). The convex set $\mathcal{S}^\star$ might not even have any finite-dimensional face! For instance, let $E$ be the space of Lebesgue-integrable functions on $[0, 1]$. If $R(u) = \int_0^1 |u(x)|\mathrm{d}x$, $\Phi : E \to \mathbb{R}$ is defined by $u \mapsto \int_0^1 u(x)\mathrm{d}x$, and $y = 1$, then

$$\mathcal{S}^\star = \left\{ u \in E \ : \ \int_0^1 |u(x)|\mathrm{d}x \leq 1 \quad \text{and} \quad \int_0^1 u(x)\mathrm{d}x = 1 \right\}. \tag{14}$$

It is possible to prove that such a set $\mathcal{S}^\star$ does not have any extreme point. As a consequence $\mathcal{S}^\star$ does not have any finite-dimensional face (otherwise an extreme point of the closure of a face would be an extreme point of $\mathcal{S}^\star$).

However, theorem 1 (in fact the Dubins-Klee theorem [20, 29]) asserts that, *if there is* a finite-dimensional face in $\mathcal{S}^\star$, then $C^\star$ has indeed extreme

points (and possibly extreme rays), and the convex combinations of such points generate the above-mentioned face.

As a result, it is crucial to be able to assert a priori the existence of some finite-dimensional face for $\mathcal{S}^\star$, and this is where topological arguments come into play. If $E$ is endowed with the topology of a locally convex (Hausdorff) vector space, the theorems [30, 3.3 and 3.4] which generalize the celebrated Krein-Milman theorem, state that $\mathcal{S}^\star$ *has an extreme point provided*

- $\mathcal{S}^\star$ is nonempty, convex,
- $\mathcal{S}^\star$ contains no line,
- and $\mathcal{S}^\star$ is closed, locally compact.

The last two conditions hold in particular if $\mathcal{S}^\star$ is compact. Moreover, as in corollary 1, the second condition can be ensured by considering a suitable quotient map, provided it preserves the other topological properties (*e.g.* if $\mathrm{lin}(C)$ has a topological complement).

**Remark 8.** *Whereas local compactness is a very strong property for topological vector spaces (implying their finite-dimensionality, see [8, Th. 3,Ch. 1]), it is not so difficult to ensure the local compactness of $\mathcal{S}^\star$ in practice. Indeed, very often, even the existence of solutions is usually ensured using compactness arguments for a suitable weak or weak-* topology. The unbounded cases require more specific arguments, but let mention that there are examples of cones which are locally compact without being contained in any finite-dimensional vector space. In Section 4.2.2 below, we discuss the example of the cone $\mathcal{M}^+(\Omega)$ of non-negative measures over a compact set for the weak-* topology. Another example of locally compact convex cone is*

$$\mathcal{C} = \left\{ x \in \mathbb{R}^{\mathbb{N}} \text{ such that } x_n \geq 0 \text{ and } \sum_{n\in\mathbb{N}} x_n \omega_n \leq \sum_{n\in\mathbb{N}} x_n < +\infty \right\} \subseteq \ell^1(\mathbb{N}),$$

*for some non-decreasing positive sequence $(\omega_n)_n$ converging to $+\infty$ (note that $\omega_0 < 1$ for the cone to be non-empty). The cone $\mathcal{C}$ is locally compact for the strong topology. Indeed, consider the intersection $K$ of the cone $\mathcal{C}$ and the strong unit ball, namely $K = \{x = (x_n)_n \,:\, x_n \geq 0,\ \sum_n \omega_n x_n \leq \sum_n x_n \leq 1\}$ and consider a sequence of elements of $K$ denoted $(x^k)_k \subset K$. Using a diagonal argument, each $(x_n^k)_k$ converges to some $\bar{x}_n \geq 0$. Furthermore, using that $\{n \,:\, w_n < 1\}$ is*

*finite and Fatou's lemma, it holds that*

$$\sum_n \bar{x}_n \leq \liminf_k \sum_n x_n^k \leq 1 \quad \text{and}$$
$$\sum_{n\,:\,w_n\geq 1} (w_n - 1)\bar{x}_n \leq \liminf_k \sum_{n\,:\,w_n\geq 1} (w_n - 1)x_n^k$$
$$\leq \liminf_k \sum_{n\,:\,w_n<1} (1 - w_n)x_n^k = \sum_{n\,:\,w_n<1} (1 - w_n)\bar{x}_n$$

*and we deduce that* $\bar{x} = (\bar{x}_n)_n \in K$. *Furthermore, one has*

$$||x^k - \bar{x}||_1 = \sum_{n=0}^{M} |x_n^k - \bar{x}_n| + \sum_{n>M} |x_n^k - \bar{x}_n|$$

*and* $M$ *will be chosen later. Finally,*

$$\sum_{n>M} |x_n^k - \bar{x}_n| \leq \sum_{n>M} (x_n^k + \bar{x}_n) \leq (1/\omega_M) \sum_{n>M} \omega_n (x_n^k + \bar{x}_n) \leq 2/\omega_M$$

*since* $\omega_n/\omega_M \geq 1$ *for* $n > M$. *Therefore choosing* $M$ *large enough ensures that the second term* $\sum_{n>M} |x_n^k - \bar{x}_n|$ *is less than some* $\varepsilon > 0$. *Choosing* $k$ *large enough leads to* $||x^k - \bar{x}||_1 \leq 2\varepsilon$.

# 4 Application to some popular regularizers

We now show that the extreme points and extreme rays of numerous convex regularizers can be described analytically, allowing to describe important analytical properties of the solutions of some popular problems. The list given below is far from being exhaustive, but it gives a taste of the diversity of applications targeted by our main results.

## 4.1 Finite-dimensional examples

We first consider examples where one has $\dim E < +\infty$. In that case, $\Phi$ is continuous, and since the considered regularizations $R$ are lower semi-continuous, we deduce that

- the level set $C^\star$ is closed,
- the solution set $\mathcal{S}^\star = C^\star \cap \Phi^{-1}(\{y\})$ is closed, and locally compact (even compact in most cases), hence it admits extreme points provided it contains no line.

### 4.1.1 Nonnegativity constraints

In a large number of applications, the signals to recover are known to be nonnegative. In that case, one may be interested in solving nonnegatively constrained problems of the form:

$$\inf_{u \in \mathbb{R}^n_+} f(\Phi u - y). \tag{15}$$

An important instance of this class of problems is the nonnegative least squares [33], which finds it motivation in a large number of applications. Applying the results of Section 3 to Problem (15) yields the following result.

**Proposition 1.** *If the solution set of* (15) *is nonempty, then it contains a solution which is $m$-sparse. In addition if $f$ is convex and the solution set is compact, then its extreme points are $m$-sparse.*

Choosing $R$ as the characteristic function of $\mathbb{R}^n_+$, the result simply stems from the fact that the extreme rays of the positive orthant are the half lines $\{\alpha e_i, \alpha \geq 0\}$, where $(e_i)_{1 \leq i \leq n}$ denote the elements of the canonical basis. We have to consider $m$ atoms and not $m - 1$ since $t^\star = \inf R = 0$, see remark 2.

It may come as a surprise to some readers, since the usual way to promote sparsity consists in using $\ell^1$-norms. This type of result is one of the main ingredients of [18] which shows that the $\ell^1$-norm can sometimes be replaced by the indicator of the positive orthant when sparse *positive* signals are looked after.

### 4.1.2 Linear programming

Let $\psi \in \mathbb{R}^n$ be a vector and $\Phi \in \mathbb{R}^{m \times n}$ be a matrix and consider the following linear program in standard (or equational) form:

$$\inf_{\substack{u \in \mathbb{R}^n_+ \\ \Phi u = y}} \langle \psi, u \rangle \tag{16}$$

Applying Theorem (1) to the problem (16), we get the following well-known result (see e.g. [34, Thm. 4.2.3]):

**Proposition 2.** *Assume that the solution set of* (16) *is nonempty and compact. Then, its extreme points are $m$-sparse, i.e. of the form:*

$$u = \sum_{i=1}^{m} \alpha_i e_i, \alpha_i \geq 0, \tag{17}$$

*where $e_i$ denotes the $i$-th element of the canonical basis.*

In the linear programming literature, solutions of this kind are called *basic solutions*. To prove the result, we can reformulate (16) as follows:

$$\inf_{\substack{(u,t)\in\mathbb{R}^n_+\times\mathbb{R}\\ \Phi u=y\\ \langle\psi,u\rangle=t}} t \tag{18}$$

Letting $R(u,t) = t + \iota_{\mathbb{R}^n_+}(u)$, we get $\inf R = -\infty$. Hence, if a solution exists, we only need to analyze the extreme points and extreme rays of $C^\star = \{(x,t) \in \mathbb{R}^n \times \mathbb{R}, R(x,t) \leq t^\star\} = \mathbb{R}^n_+ \times ]-\infty, t^\star]$, where $t^\star$ denotes the optimal value. The extreme rays of this set (a shifted nonnegative orthant) are of the form $\{\alpha e_i, \alpha > 0\} \times \{t^\star\}$ or $\{0\}\times]-\infty, t^\star[$. In addition $C^\star$ possesses only one extreme point $(0, t^\star)$. Applying theorem 1, we get the desired result.

### 4.1.3 $\ell^1$ analysis priors

An important class of regularizers in the finite dimensional setting $E = \mathbb{R}^n$ contains the functions of the form $R(u) = \|Lu\|_1$, where $L$ is a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^p$. They are sometimes called *analysis priors*, since the signal $u$ is "analyzed" through the operator $L$. Remarkable practical and theoretical results have been obtained using this prior in the fields of inverse problems and compressed sensing, even though many of its properties are -to the belief of the authors- still quite obscure.

Since $R$ is one-homogeneous, it suffices to describe the extremality properties of the unit ball $C = \{u \in \mathbb{R}^n, \|Lu\|_1 \leq 1\}$ to use our theorems. The lineality space is simply equal to $\mathrm{lin}(C) = \ker(L)$. Let $K = \ker(L)$, $K^\perp$ denote the orthogonal complement of $K$ in $\mathbb{R}^n$ and $L^+ : \mathbb{R}^n \to K^\perp$ denote the pseudo-inverse of $L$. We can decompose $C$ as $C = K + C_{K^\perp}$ with $C_{K^\perp} = C \cap K^\perp$. Our ability to characterize the extreme points of $C_{K^\perp}$ depends on whether $L$ is surjective or not. Indeed, we have

$$\mathrm{ext}(C_{K^\perp}) = L^+ \left(\mathrm{ext}\left(\mathrm{ran}(L) \cap B_1^p\right)\right), \tag{19}$$

where $B_1^p$ is the unit $\ell^1$-ball defined as

$$B_1^p = \{z \in \mathbb{R}^p, \|z\|_1 \leq 1\}.$$

Property (19) simply stems from the fact that $C_{K^\perp}$ and $D = \mathrm{ran}(L) \cap B_1^p$ are in bijection through the operators $L$ and $L^+$.

**The case of a surjective operator $L$** When $L$ is surjective (hence $p \leq n$), the problem becomes quite elementary.

**Proposition 3.** *If $L$ is surjective, the extreme points $u$ of $C_{K^\perp}$ are $\mathrm{ext}(C_{K^\perp}) = (\pm L^+ e_i)_{1\leq i\leq p}$, where $e_i$ denotes the $i$-th element of the canonical basis. Consider Problem ($\mathcal{P}_f$) and assume that at least one solution exists. Then Problem ($\mathcal{P}_f$) has solutions of the form*

$$u^\star = \sum_{i\in I} \alpha_i L^+ e_i + u_K, \tag{20}$$

*where $u_K \in \ker(L)$ and $I \subset \{1,\dots,p\}$ is a set of cardinality $|I| \leq m - \dim(\Phi\ker(L))$.*

The proof of Proposition 3 follows from Corollary 1 and Section 3.3.2, with $j = 0$ and observing that $\pi_K$ is the orthogonal projection on $K^\perp$.

**The case of an arbitrary operator $L$** When $L$ is not surjective the description of the extreme points $\mathrm{ext}\,(D)$ becomes untractable in general. A rough upper-bound on the number of extreme points can be obtained as follows. We assume that $L$ has full rank $n$ and that $\mathrm{ran}(L)$ is in general position. The extreme points of $\mathrm{ran}(L) \cap B_1^p$ correspond to the intersections of some faces of the $\ell^1$-ball with a subspace of dimension $n$. In order for some $k$-face to intersect the subspace $\mathrm{ran}(L)$ on a singleton, $k$ should satisfy $n + k - p = 0$, i.e. $k = p - n$. The $k$-faces of the $\ell^1$-ball contain $(k+1)$-sparse elements. The number of $(k+1)$-sparse supports in dimension $p$ is $\binom{p}{k+1}$. For a fixed support, the number of sign patterns is upper-bounded by $2^{k+1}$. Hence, the maximal number of extreme points satisfies $|\,\mathrm{ext}\,(\mathrm{ran}(L) \cap B_1^p)\,| \leq 2^{k+1}\binom{p}{k+1}$. This upper-bound is pessimistic since the subspace may not cross all extreme points, but it provides an idea of the combinatorial explosions that may happen in general.

Notice that enumerating the number of faces of a polytopes is usually a hard problem. For instance, the Motzkin conjecture [36] which upper bounds the number of $k$ faces of a $d$ polytope with $z$ vertices was formulated in 1957 and solved by Mc Mullen [35] in 1970 only.

### 4.1.4 Matrix examples

In several applications, one deals with optimization problems in matrix spaces. The following regularizations/convex sets are commonly used.

**Semi-definite matrix constraint** Similarly to Section 4.1.1, one may consider in $\mathbb{R}^{n\times n}$ the following constrained problem

$$\inf_{M\succeq 0} f(\Phi M - y), \tag{21}$$

where $M \succeq 0$ means that $M$ must be symmetric positive semi-definite (p.s.d.). The extreme rays of the positive semi-definite cone $C^\star$ are the p.s.d. matrices of rank 1 (see for instance [15, Sec. 2.9.2.7]). Hence, arguing as in Section 4.1.1, we may deduce that if there exists a solution to (21), there is also a solution which has rank (at most) $m$.

However, that conclusion is not optimal, as in that case a theorem by Barvinok [5, Th. 2.2] ensures that there exists a solution $M$ with

$$\operatorname{rank}(M) \leq \frac{1}{2}\left(\sqrt{8m+1}-1\right). \tag{22}$$

To understand the gap with Barvinok's result, let us note that the p.s.d. cone has a very special structure which makes the Minkowski-Carathéodory theorem (or its extension by Klee) too pessimistic. By [15, Sec. 2.9.2.3], given $M \succeq 0$, the smallest face of the p.s.d. cone which contains $M$ (*i.e.* the set of p.s.d matrices which have the same kernel) has dimension

$$d = \frac{1}{2}\operatorname{rank}(M)(\operatorname{rank}(M)+1). \tag{23}$$

Equivalently, if the smallest face which contains $M$ has dimension $d$, then $\operatorname{rank}(M) = \frac{1}{2}\left(\sqrt{8d+1}-1\right)$, hence $M$ is a convex combination of $\frac{1}{2}\left(\sqrt{8d+1}-1\right)$ points in extreme rays, a value which is less than the value $d$ predicted by Klee's extension of Carathéodory's theorem.

As a result, we recover Barvinok's result by noting that, as ensured by the first claim[3] of theorem 1, any extreme point $M$ of the solution set belongs to a face of dimension $m$. Then, taking into account (23) improves upon the second claim of theorem 1, and we immediately obtain (22).

**Semi-definite programming** Semi-definite programs are problems of the form:

$$\inf_{\substack{M\succeq 0\\ \Phi(M)=y}} \langle A, M\rangle, \tag{24}$$

where $A \in \mathbb{R}^{n\times n}$ is a matrix and $\langle A, M\rangle \overset{\text{def.}}{=} \operatorname{Tr}(AM)$. Arguing as in proposition 2, if the solution set of (24) is nonempty, our main result allows to state that its extreme points are matrices of rank $m$. In view of the above discussion, it is possible to refine this statement and show that (22) holds.

[3] or, more precisely, by its variant when $t^\star = \inf R$, see remark 2.

**The nuclear norm** The nuclear norm of a matrix $M \in \mathbb{R}^{p\times n}$ is often denoted $\|M\|_*$ and defined as the sum of the singular values of $M$. It gained a considerable attention lately as a regularizer thanks to its applications in matrix completion [11] or blind inverse problems [2]. The geometry of the unit ball $\{M \in \mathbb{R}^{p\times n}, \|M\|_* \leq 1\}$ is well studied due to its central role in the field of semi-definite programming [38]. Its extreme points are the rank one matrices $M = uv^T$, with $\|u\|_2 = \|v\|_2 = 1$.

Combining theorem 1 with this result explains why regularizing problems over the space of matrices with the nuclear norm allows recovering *rank-m solutions*.

**The rank-sparsity ball** The *rank-sparsity* ball is the set $\{M \in \mathbb{R}^{m\times n}, \|M\|_* + \|M\|_1 \leq 1\}$, where $\|M\|_1$ is the $\ell^1$-norm of the entries of $M$. The corresponding regularization is sometimes used in order to favor sparse and low-rank matrices. The authors of [19] have described the extreme points of this unit ball. They have proved that *the extreme points $M$ of the rank sparsity ball satisfy* $\frac{r(r+1)}{2} - |I| \leq 1$, where $|I|$ denotes the number of non-zero entries in $M$ and $r$ denotes its rank. This result partly explains why using the rank-sparsity gauge promotes sparse and low rank solution. Let us outline that this effect might be better obtained using different strategies [39, 40].

**Bi-stochastic matrices** A doubly stochastic matrix is a matrix with non-negative rows and columns summing to one. The set of such matrices is called the Birkhoff polytope. The Birkhoff-von Neumann theorem states that its extreme points are the permutation matrices. We refer the interested reader to [26] for an use of such matrices in DNA sequencing.

## 4.2 Examples in infinite dimension

In this section, we provide results in infinite dimensional spaces, which echoe the ones described in finite dimension.

### 4.2.1 Problems formulated in Hilbert or Banach sequence spaces

The case of Hilbert spaces (or countable sequences) can be treated within our formalism and all the examples presented previously have their natural counterpart in this setting. In the same vein, one can also treat Banach sequence spaces $\ell^p$ for $1 \leq p \leq \infty$. We do not reproduce the results here

for space limitations. Let us however mention that two works treat this specific case with $\ell^1$ regularizers [47, 1].

### 4.2.2 Linear programming and the moment problem

Let $\Omega$ be a compact metric space, $\mathcal{M}(\Omega)$ be the set of Radon measures on $\Omega$ and let $\mathcal{M}_+(\Omega) \subseteq \mathcal{M}(\Omega)$ be the cone of nonnegative measures on $\Omega$. Let $\psi$ and $(\phi_i)_{1\leq i\leq m}$ denote a collection of continuous functions on $\Omega$. Now, let $\Phi : \mathcal{M}(\Omega) \to \mathbb{R}^m$ be defined by $(\Phi\mu)_i = \langle \phi_i, \mu\rangle$, where $\langle \phi_i, \mu\rangle \overset{\text{def.}}{=} \int_\Omega \phi_i \mathrm{d}\mu$, and consider the following linear program in standard form:

$$\inf_{\substack{\mu\in\mathcal{M}_+(\Omega)\\ \Phi\mu=y}} \langle \psi, u\rangle. \tag{25}$$

Applying Theorem (1) to the problem (25), we get proposition 4 below. We do not provide a proof here, since it mimics very closely the one given for linear programming in finite dimension. The extreme rays of $\mathcal{M}_+(\Omega)$ can be described, arguing as in [3, Th. 15.9], as the rays directed by the Dirac masses.

**Proposition 4.** *Assume that the solution set* (25) *is nonempty. Then, its extreme points are $m$-sparse, i.e. of the form:*

$$u = \sum_{i=1}^{m} \alpha_i \delta_{x_i}, x_i \in \Omega, \alpha_i \geq 0. \tag{26}$$

To make sure that the above proposition is non-trivial, one may wish to ensure that the solution set $\mathcal{S}^\star$ has indeed extreme points, using arguments from Section 3.4. It is straightforward that $\mathcal{S}^\star$ is convex and does not contain any line. Now, let us endow $\mathcal{M}(\Omega)$ with the weak-* topology (*i.e.* the coarsest topology for which $\mu \mapsto \int_\Omega \eta \mathrm{d}\mu$ is continuous for every $\eta \in \mathscr{C}(\Omega)$). By lower semi-continuity, $\mathcal{S}^\star$ is closed. Moreover, $\mathcal{S}^\star$ is locally compact since the closed convex cone $\mathcal{M}_+$ is itself locally compact (take any $\mu \in \mathcal{M}_+(\Omega)$, its neighborhood $\{\nu \in \mathcal{M}_+(\Omega) \ : \ \nu(\Omega) \leq \mu(\Omega) + 1\}$ is compact in the weak-* topology).

Proposition 4 is well known, see e.g. [44]. Note that if we optimize the linear form $\langle \psi, u\rangle$ over the set of probability measures instead of the set of nonnegative measures, we get the so-called moment problem [45] for which we can obtain a similar result.

### 4.2.3 The total variation ball

Let $\Omega$ denote an open subset of $\mathbb{R}^d$ and $\mathcal{M}(\Omega)$ denote the set of Radon measures on $\Omega$. The total variation ball $B_{\mathcal{M}} = \{u \in \mathcal{M}(\Omega), \|u\|_{\mathcal{M}(\Omega)} \leq 1\}$ plays a critical role for problems such as super-resolution [10, 46, 21]. It is compact for the weak-* topology and its extreme points are the Dirac masses: $\mathrm{ext}(B_{\mathcal{M}}) = \{\pm\delta_x, x \in \Omega\}$. Hence total variation regularized problems of the form:

$$\inf_{u \in \mathcal{M}} f(\Phi u) + \|u\|_{\mathcal{M}},$$

yield $m$-sparse solutions (under an existence assumption). A few variations around this central result were provided in [22].

**Demixing of Sines and Spikes** In [22, Page 262], the author presents a regularization of the type

$$\|\mu\|_{\mathcal{M}} + \eta\|v\|_1$$

where $\eta > 0$ is a tuning parameter, $\mu$ a complex measure and $v \in \mathbb{C}^n$ a sparse vector. Define $E$ as the set of $(\mu, v)$ where $\mu$ is a complex Radon measure on a domain $\Omega$ and $v \in \mathbb{C}^n$. Consider the unit ball

$$B \stackrel{\text{def.}}{=} \{(\mu, v) \in E \ : \ \|\mu\|_{\mathcal{M}} + \eta\|v\|_1 \leq 1\}.$$

Its extreme points are the points

- $(a\delta_t, 0)$ for all $t \in \Omega$ (and $\delta_t$ denotes the Dirac mass at point $t$) and all $a \in \mathbb{C}$ such that $|a| = 1$,
- $(0, ae_k)$ for all $k = 1, \ldots, n$ and all $a \in \mathbb{C}$ such that $|a| = 1/\eta$ and $e_k$ denotes the vector with $1$ at entry $k$ and $0$ otherwise.

**Group Total Variation: Point sources with a common support** In [22, Page 266], the author presents a regularization of the type

$$\|\mu\|_{\mathcal{M}^n} := \sup_{F:\Omega\to\mathbb{C}^n,\ \|F(t)\|_2\leq 1,\ t\in\Omega} \int_\Omega \langle F(t), \nu(t)\rangle \mathrm{d}|\mu|(t)$$

were $F$ is continuous and vanishing at infinity, and $\mu$ is a vectorial Radon measure on $\Omega$ such that $|\mu|$-a.e. $\mu = \nu \cdot |\mu|$ with $\nu$ a measurable function from $\Omega$ onto $\mathbb{S}^{n-1}$ the $n$-sphere and $|\mu|$ a positive finite measure on $\Omega$. Consider the unit ball

$$B \stackrel{\text{def.}}{=} \{\mu, \|\mu\|_{\mathcal{M}^n} \leq 1\}.$$

Its extreme points are $a\delta_t$ for all $t \in \Omega$ (and $\delta_t$ denotes the Dirac mass at point $t$) and all $a \in \mathbb{C}^n$ such that $\|a\|_2 = 1$.

### 4.2.4 Analysis priors in Banach spaces

The analysis of extreme points of analysis priors in an infinite dimensional setting is more technical. Fisher and Jerome [23] proposed an interesting result, which can be seen as an extension of (20). This result was recently revisited in [48] and [25]. Below, we follow the presentation in [25].

Let $\Omega$ denote an open set in $\mathbb{R}^d$. Let $\mathcal{D}'(\Omega)$ denote the set of distributions on $\Omega$ and let $L : \mathcal{D}'(\Omega) \to \mathcal{D}'(\Omega)$ denote a linear operator with kernel $K = \ker(L)$. We let $E = \{u \in \mathcal{D}'(\Omega), Lu \in \mathcal{M}(\Omega)\}$ and let $\|\cdot\|_K$ denote a semi-norm on $E$, which restricted to $K$ is a norm. We define a function space $\mathcal{B}(\Omega)$ as follows:

$$\mathcal{B}(\Omega) = \{u \in E, \|Lu\|_{\mathcal{M}(\Omega)} + \|u\|_K < +\infty\}$$

and equip it with the norm $\|u\|_{\mathcal{B}(\Omega)} = \|Lu\|_{\mathcal{M}(\Omega)} + \|u\|_K$. We assume that $L$ is surjective, i.e. $\mathcal{M}(\Omega) = L(\mathcal{B}(\Omega))$, and that $K$ has a topological complement (with respect to $\mathcal{B}(\Omega)$), which we denote by $K^\perp$. This setting encompasses all surjective Fredholm operators for instance. Under the stated assumptions, we can define a pseudo-inverse $L^+$ of $L$ relative to $K^\perp$ [7].

The representer theorems in [23, 48, 25] can be obtained using theorem 1 as exemplified below.

**Proposition 5.** *Let $B = \{u \in \mathcal{B}(\Omega), \|Lu\|_{\mathcal{M}(\Omega)} \leq 1\}$. Then the extreme points of the set $C_{K^\perp} = B \cap K^\perp$ are of the form $\pm L^+\delta_x$, for $x \in \Omega$.*

*Let $f : \mathbb{R}^m \to \mathbb{R} \cup \{+\infty\}$ denote a convex function and define*

$$\mathcal{S}^\star = \operatorname{argmin}_{u \in \mathcal{B}(\Omega)} f(\Phi u) + \|Lu\|_{\mathcal{M}(\Omega)}.$$

*Assume that $\mathcal{S}^\star$ is nonempty and does not contain $0$. Then the extreme points (if they exist) of $\pi_K(\mathcal{S}^\star)$ are of the form $u = \sum_{i=1}^m \alpha_i L^+\delta_{x_i}$.*

*Proof.* The proof mimics the finite dimensional case (20). First notice that $B = L^{-1}(B_{\mathcal{M}})$, where $L^{-1}(\{\mu\})$ is the pre-image of $\mu$ by $L$ and $B_{\mathcal{M}}$ is the unit total variation ball. We have $L^{-1}(B_{\mathcal{M}}) = L^+(B_{\mathcal{M}}) + K$ and we can identify $C_{K^\perp}$ with $L^+(B_{\mathcal{M}})$. Since $L^+$ is bijective from $\mathcal{M}(\Omega)$ to $K^\perp$, the extreme points of $C_{K^\perp}$ are the image by $L^+$ of the Dirac masses.

The end of the proposition follows from Corollary 2 and from the fact that the lineality space of $\{u \in \mathcal{B}(\Omega), \|Lu\|_{\mathcal{M}(\Omega)} \leq 1\}$ is equal to $K$. □

Let us mention that, although the description of the extreme points follows directly from the results of Section 3, proving the existence of minimizers and the existence of extreme points is a considerably more difficult problem which needs a careful choice of topologies. The paper [48] provides a systematic way to construct Banach spaces and pseudo-inverse $L^+$

for "spline admissible operators" $L$ such as the fractional Laplacian. In addition, they prove existence of solutions by adding weak-* continuity assumptions on the sensing operator $\Phi$.

### 4.2.5 The total gradient variation

Since its introduction in the field of image processing [41], the total gradient variation proved to be an extremely valuable regularizer in diverse fields of data science and engineering. It is defined, for any locally integrable function $u$ as

$$TV(u) \overset{\text{def.}}{=} \sup \left( \int u \operatorname{div}(\phi)\, dx, \phi \in C_c^1(\mathbb{R}^d)^d, \sup_{x \in \mathbb{R}^d} \|\phi(x)\|_2 \leq 1 \right).$$

If the above quantity is finite, we say that $u$ has bounded variation and its gradient $Du$ is a Radon measure, with

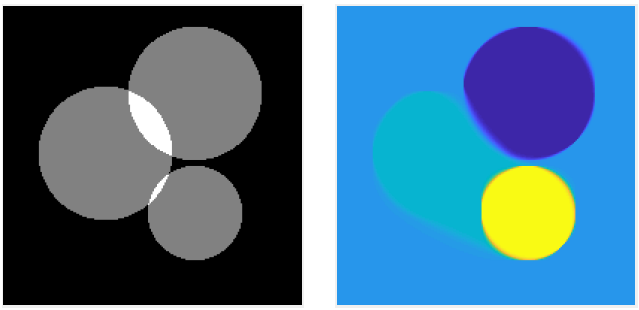
$$TV(u) = \int_{\mathbb{R}^d} |Du| = \|Du\|_{(\mathcal{M}(\mathbb{R}^d))^d}.$$

Working in $E = L^{d/(d-1)}(\mathbb{R}^d)$, one is led to consider the convex set $C = \{u \in E, TV(u) \leq 1\}$, referred to as the TV unit ball.

The generalized gradient operator is not a surjective operator. Hence, the analysis of Section 4.2.4 cannot help finding the extreme points of the TV ball. Still, those have been described in the fifties by Fleming in [24] and refined analyses have been proposed more recently by Ambrosio, Caselles, Masnou and Morel in [4].

**Theorem 2** (Extreme points of the TV ball [24, 4])**.** *The extreme points of the unit TV unit ball are the indicators of simple sets normalized by their perimeter, i.e. functions of the form $u = \pm \frac{\mathbb{1}_F}{TV(\mathbb{1}_F)}$, where $F$ is an indecomposable and saturated subset of $\mathbb{R}^d$.*

Informally, the simple sets of $\mathbb{R}^d$ are the simply connected sets with no hole. We refer the reader to [4] for more details. Using theorem 2 in conjunction with our results tell us that functions minimizing the total variation subject to a finite number of linear constraints can be expressed as a sum of a small number of indicators of simple sets, see for instance fig. 3, which is yet another theoretical result explaining the common observation that total variation tends to produce stair-casing [37].



(a) (b)

Figure 3: Illustration for the total gradient variation problem $\min\{TV(u) : \Phi(u) = y\}$. Here, $\Phi$ is a linear mapping giving access to 3 measurements, $y \in \mathbb{R}^3$, by performing the mean of an image $u$ of size $200 \times 200$ on 3 different disks represented in (a). The TV problem is solved using a primal-dual algorithm, also known as the Chambolle-Pock algorithm [12]. The recovered image is displayed in (b): it can be represented as the sum of 3 indicator functions of simple sets.

# 5 Conclusion

In this paper we have developed representer theorems for convex regularized inverse problems (1), based on fundamental properties of the geometry of convex sets: the solution set can be entirely described using convex combinations of a small number of extreme points and extreme rays of the regularizer level set.

Obviously, the conclusion of Theorem 1 is only nontrivial when $C^\star$ has a "sufficiently flat boundary", in the sense that two or more faces of $C^\star$ have dimension larger than $m$. For instance, if $C^\star$ is strictly convex (*i.e.* has only 0-dimensional faces, except its interior[4]), then the solution set $\mathcal{S}^\star$ is always reduced to a single extreme point of $C^\star$! Nevertheless, several regularizers which are commonly used in the literature (notably sparsity-promoting ones) have that flatness property, and Theorem 1 then provides interesting information on the structure of the solution set, as illustrated in Section 4.

To conclude, the structure theorem presented in this paper highlights the importance of describing the extreme points and extreme rays of the regularizer: this yields a fine description of the set of solutions of variational problems of the form (1). Our theorem also suggests a principled

[4]In this example, to simplify the discussion, we assume that $E$ has finite dimension.

way to design a regularizer. If a particular family of solutions is expected, then one may construct a suitable regularizer by taking the convex hull of this family. Finally, representer theorems have had a lot of success in the fields of approximation theory and machine learning [42], in the frame of reproducible kernel Hilbert spaces. A reason of this success is that they allow to design efficient numerical procedures that yield *infinite dimensional* solutions by solving *finite dimensional* linear systems. Such numerical procedures have recently been extended to the case of Banach spaces for some simple instances of the problems described in this paper [22, 16, 25]. The price to pay when going from a Hilbert space to a Banach space is that semi-infinite convex programs have to be solved instead of simpler linear systems. We foresee that the results in this paper may help designing new efficient numerical procedures, since they allow to parameterize the solutions using extreme points and extreme rays only.

# 6 Proofs of Section 3

## 6.1 Proof of Theorem 1

The set of solutions $\mathcal{S}^\star$ is precisely $C^\star \cap \Phi^{-1}(\{y\})$, and the statement of the theorem amounts to describing its elementary faces. Since $\Phi^{-1}(\{y\})$ is an affine space with codimension at most $m$, the main theorem of [29] almost provides the desired conclusion, but, for our particular case, it yields one extreme point/ray too many. Here is how to obtain the correct number.

Let $p$ be a point of $\mathcal{S}^\star$ such that $\mathcal{F}_{\mathcal{S}^\star}(p)$ has dimension $j$. Up to a translation, it is not restrictive to assume that $p = 0$, so that $\mathcal{S}^\star = C^\star \cap \ker\Phi$.

Let $T$ be the union of $\{0\}$ and all the lines $\ell$ such that $C^\star \cap \ell$ contains an open interval which contains $0$. Note that $T$ is a linear space, the linear hull of $\mathcal{F}_{C^\star}(0)$.

We claim that $\operatorname{codim}_T (T \cap \ker\Phi) \leq m-1$. By contradiction assume that there is a complement $Z$ to $T \cap \ker\Phi$ in $T$ with dimension $m$. Then $\Phi|_Z$ has rank $m$ and is a bijection, hence we may define

$$z \stackrel{\text{def.}}{=} -\frac{\theta}{(1-\theta)}(\Phi|_Z)^{-1}(\Phi u_0) \in Z \subset T,$$

where $\theta \in ]0,1[$ and $u_0 \in C^\star$ is such that $\inf R \leq R(u_0) < t^\star$. For $\theta$ small enough, $z \in C^\star$, hence $R(z) \leq t^\star$. Moreover,

$$\Phi\left((1-\theta)z + \theta u_0\right) = (1-\theta)\Phi z + \theta\Phi u_0 = 0. \tag{27}$$

so that $(1-\theta)z+\theta u_0$ lies in $C^\star \cap \ker\Phi$. Since $R(u_0) < R(z)$, and $R$ is convex

$$R\left((1-\theta)z+\theta u_0\right) < R\left(z\right) \leq t^\star, \tag{28}$$

we obtain a contradiction with the fact that $t^\star$ is the minimal value of ($\mathcal{P}$). As a result, $\mathrm{codim}_T\left(T\cap\ker\Phi\right) \leq m-1$. Observing that $T\cap\ker\Phi$ is the linear hull of $\mathcal{F}_{\mathcal{S}^\star}(0)$, hence $j = \dim\left(T\cap\ker\Phi\right)$, we deduce that

$$\dim\mathcal{F}_{C^\star}(0) \overset{\text{def.}}{=} \dim T = \mathrm{codim}_T\left(T\cap\ker\Phi\right)+\dim\left(T\cap\ker\Phi\right) \leq m-1+j, \tag{29}$$

and the first claim of the theorem is proved.

Now, applying the Carathéodory-Klee theorem (3) in [29], $p$ is convex combination of at most $m+j$ (resp. $m+j-1$) extreme points (resp. extreme points or in an extreme ray) of $\mathcal{F}_{C^\star}(0)$. The conclusion stems from the fact that the extreme points (resp. rays) of $\mathcal{F}_{C^\star}(0)$ are extreme points (resp. rays) of $C^\star$, see the proof of the main theorem in [29].

□

## 6.2 Proof of Corollary 1

Now, assume that the vector space $K \overset{\text{def.}}{=} \mathrm{rec}(C^\star)\cap(-\mathrm{rec}(C^\star))$ is non trivial (otherwise the conclusion follows from Theorem 1). We note that for any $u\in C^\star$, the convex function $v\mapsto R(u+v)$ is upper bounded by $t^\star$ on $K$, hence is constant. As a result, possibly replacing $R$ with $R+\chi_{C^\star}$, it is not restrictive to assume that $R$ is invariant by translation along $K$.

Now, let $\pi_K$, $\pi_{\Phi(K)}$ be the canonical quotient maps and define $\tilde{R}$ and $\tilde{\Phi}$ by the commutative diagrams

$$\begin{array}{ccc} E & \xrightarrow{R} & [-\infty,\infty] \\ {\scriptstyle\pi_K}\downarrow & \nearrow_{\tilde{R}} & \\ E/K & & \end{array} \qquad \begin{array}{ccc} E & \xrightarrow{\Phi} & \mathbb{R}^m \\ {\scriptstyle\pi_K}\downarrow & & \downarrow{\scriptstyle\pi_{\Phi(K)}} \\ E/K & \xrightarrow{\tilde{\Phi}} & \mathbb{R}^m/\Phi(K) \end{array}$$

Note that $\tilde{R}$ is a convex function and that $\tilde{\Phi}$ is a linear map with rank $m-d$, where $d \overset{\text{def.}}{=} \dim\left(\Phi(K)\right)$. It is then natural to consider the problem

$$\min_{\tilde{u}\in E/K} \tilde{R}(\tilde{u}) \quad \text{s.t.} \quad \tilde{\Phi}\tilde{u} = \tilde{y}, \tag{$\tilde{\mathcal{P}}$}$$

where $\tilde{y} \overset{\text{def.}}{=} \pi_{\Phi(K)}(y)$. In other words, one still wishes to minimize $R(u)$, but one is satisfied if the constraint $\Phi u = y$ merely holds up to an additional

term $\Phi v$, where $v \in K$. We observe that ($\mathcal{P}$) and ($\tilde{\mathcal{P}}$) have the same value $t^\star$, and the level set

$$\tilde{C}^\star \overset{\text{def.}}{=} \left\{ \tilde{u} \in E/K \ : \ \tilde{R}(\tilde{u}) \leq t^\star \right\} = \pi_K(C^\star) \tag{30}$$

is convex linearly closed *and contains no line*. Let $\tilde{\mathcal{S}}^\star$ be the solution set to ($\tilde{\mathcal{P}}$). Theorem 1 now describes the elements of the $j$-dimensional faces of $\tilde{\mathcal{S}}^\star$ as convex combinations of $m - d + j$ (resp. $m - d + j - 1$) extreme points (resp. extreme points or points in an extreme ray) of $\tilde{C}^\star$ that we denote by $\tilde{q}_1$,$\tilde{q}_2$,$\ldots$, $\tilde{q}_r$.

To conclude, we have obtained $\pi_K(p) = \sum_i \theta_i \tilde{q}_i$, for some $\theta \in \mathbb{R}_+^r$ with $\sum_i \theta_i = 1$. Equivalently, since $\psi_K^{-1}(\cdot, 0)$ provides one element in the corresponding class, this means that $p \in \psi_K^{-1}(\sum_i \theta_i \tilde{q}_i, 0) + K$. We get the claimed result by linearity of $\psi_K^{-1}$. □

**Remark 9.** *Incidentally, we note that for $Y = \{y\}$, the face $\mathcal{F}_{\mathcal{S}^\star_{\{y\}}}(p)$ is isomorphic (through $\psi_K$) to $\mathcal{F}_{\pi_K(\mathcal{S}^\star_{\{y\}})}(\pi_K(p)) \times (K \cap \ker\Phi)$.*

# Acknowledgments

This work was initially started by two different groups composed of A. Chambolle, F. de Gournay and P. Weiss on one side and C. Boyer, Y. De Castro and V. Duval on the other side. The authors realized that they were working on a similar topic when a few of them met at the Cambridge semester in mathematical imaging during the Isaac Newton Institute (Cambridge) semester "Variational methods and effective algorithms for imaging and vision", supported by EPSRC Grant N. EP/K032208/1. They therefore decided to write a joint paper and wish to thank the organizers for giving them this opportunity. This work initially started with the help of T. Pock through a few numerical experiments and with discussions with J. Fadili and C. Poon.

The work of A.C. was partially supported by a grant of the Simons Foundation.

# References

[1] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Foundations of Computational Mathematics*, 16(5):1263–1323, 2016.

[2] A. Ahmed, B. Recht, and J. Romberg. Blind deconvolution using convex programming. *IEEE Transactions on Information Theory*, 60(3):1711–1732, 2014.

[3] C. Aliprantis and K. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer Berlin Heidelberg, 2006.

[4] L. Ambrosio, V. Caselles, S. Masnou, and J.-M. Morel. Connected components of sets of finite perimeter and applications to image processing. *Journal of the European Mathematical Society*, 3(1):39–92, 2001.

[5] A. I. Barvinok. Problems of distance geometry and convex properties of quadratic maps. *Discrete & Computational Geometry*, 13(2):189–202, Mar. 1995.

[6] A. Beurling. Sur les intégrales de Fourier absolument convergentes et leur application à une transformation fonctionnelle. In *Ninth Scandinavian Mathematical Congress*, pages 345–366, 1938.

[7] F. J. Beutler. The operator theory of the pseudo-inverse i. bounded operators. 10, 06 1965.

[8] N. Bourbaki. *Espaces vectoriels topologiques: Chapitres 1à 5*. Springer-Verlag, Berlin Heidelberg, 2007.

[9] K. Bredies and M. Carioni. Sparsity of solutions for variational inverse problems with finite-dimensional data. *arXiv:1809.05045 [math]*, Sept. 2018. arXiv: 1809.05045.

[10] E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[11] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717, 2009.

[12] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, May 2011.

[13] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, 2012.

[14] A. Daniilidis and Y. G. Ramos. Some remarks on the class of continuous (semi-) strictly quasiconvex functions. *Journal of optimization theory and applications*, 133(1):37–48, 2007.

[15] J. Dattorro. *Convex Optimization & Euclidean Distance Geometry*. Meboo Publishing, 2005.

[16] Y. De Castro, F. Gamboa, D. Henrion, and J.-B. Lasserre. Exact solutions to super resolution on semi-algebraic domains in higher dimensions. *IEEE Transactions on Information Theory*, 63(1):621–630, 2017.

[17] D. L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[18] D. L. Donoho and J. Tanner. Sparse nonnegative solution of underdetermined linear equations by linear programming. *Proceedings of the National Academy of Sciences of the United States of America*, 102(27):9446–9451, 2005.

[19] D. Drusvyatskiy, S. A. Vavasis, and H. Wolkowicz. Extreme point inequalities and geometry of the rank sparsity ball. *Mathematical Programming*, 152(1-2):521–544, 2015.

[20] L. E. Dubins. On extreme points of convex sets. *Journal of Mathematical Analysis and Applications*, 5(2):237–244, 1962.

[21] V. Duval and G. Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.

[22] C. Fernandez-Granda. Super-resolution of point sources via convex programming. *Information and Inference: A Journal of the IMA*, 5(3):251–303, 2016.

[23] S. D. Fisher and J. W. Jerome. Spline solutions to L1 extremal problems in one and several variables. *Journal of Approximation Theory*, 13(1):73–83, 1975.

[24] W. Fleming. Functions with generalized gradient and generalized surfaces. *Annali di Matematica Pura ed Applicata*, 44(1):93–103, 1957.

[25] A. Flinth and P. Weiss. Exact solutions of infinite dimensional total-variation regularized problems. *arXiv preprint arXiv:1708.02157*, 2017.

[26] F. Fogel, R. Jenatton, F. Bach, and A. D'Aspremont. Convex relaxations for permutation problems. *Advances in Neural Information Processing Systems*, 36(4):1016–1024, 2015.

[27] H. Gupta, J. Fageot, and M. Unser. Continuous-domain solutions of linear inverse problems with tikhonov vs. generalized tv regularization. *arXiv preprint arXiv:1802.01344*, 2018.

[28] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I*, volume 305 of *Grundlehren der mathematischen Wissenschaften*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1993.

[29] V. Klee. On a theorem of Dubins. *Journal of Mathematical Analysis and Applications*, 7(3):425–427, Dec. 1963.

[30] V. L. Klee. Extremal structure of convex sets. *Archiv der Mathematik*, 8(3):234–240, Aug. 1957.

[31] M. Kreĭn. The L-problem in an abstract normed linear space. In I. Ahiezer and M. Kreĭn, editors, *Some questions in the theory of moments*, chapter 4. Gos. Naučno-Tehn. Izdat. Ukraine, 1938. English Transl. Amer. Math. Soc., Providence, R.I., 1962. MR 29 # 5073.

[32] M. G. Krein and A. A. Nudelman. *The Markov moment problem and extremal problems: ideas and problems of P. L. Cebysev and A. A. Markov and their further development*. Number v. 50 in Translations of mathematical monographs. American Mathematical Society, Providence, R.I, 1977.

[33] C. L. Lawson and R. J. Hanson. *Solving least squares problems*, volume 15. Siam, 1995.

[34] J. Matousek and B. Gärtner. *Understanding and using linear programming*. Springer Science & Business Media, 2007.

[35] P. McMullen. The maximum numbers of faces of a convex polytope. *Mathematika*, 17(2):179–184, 1970.

[36] T. S. Motzkin. Comonotone curves and polyhedra. *Bull. Amer. Math. Soc.*, 63, 1957.

[37] M. Nikolova. Local strong homogeneity of a regularized estimator. *SIAM Journal on Applied Mathematics*, 61(2):633–658, 2000.

[38] G. Pataki. The geometry of semidefinite programming. In *Handbook of semidefinite programming*, pages 29–65. Springer, 2000.

[39] E. Richard, F. Bach, and J. P. Vert. Intersecting singularities for multi-structured estimation. In *International Conference on International Conference on Machine Learning*, pages III–1157, 2013.

[40] E. Richard, G. R. Obozinski, and J.-P. Vert. Tight convex relaxations for sparse matrix factorization. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 3284–3292. Curran Associates, Inc., 2014.

[41] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.

[42] B. Schölkopf, R. Herbrich, and A. J. Smola. A generalized representer theorem. In *International conference on computational learning theory*, pages 416–426. Springer, 2001.

[43] B. Scholkopf and A. J. Smola. *Learning with kernels :*. MIT Press,, 2002.

[44] A. Shapiro. On duality theory of conic linear problems. In *Semi-infinite programming*, pages 135–165. Springer, 2001.

[45] J. A. Shohat and J. D. Tamarkin. *The problem of moments*. Number 1. American Mathematical Soc., 1943.

[46] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressed sensing off the grid. *IEEE Transactions on Information Theory*, 59(11):7465–7490, 2013.

[47] M. Unser, J. Fageot, and H. Gupta. Representer Theorems for Sparsity-Promoting $\ell_1$ Regularization. *IEEE Transactions on Information Theory*, 62(9):5167–5180, 2016.

[48] M. Unser, J. Fageot, and J. P. Ward. Splines are universal solutions of linear inverse problems with generalized tv regularization. *SIAM Review*, 59(4):769–793, 2017.

[49] H. Wendland. *Scattered data approximation*. Cambridge University Press, 2005.

[50] S. Zuhovickiĭ. Remarks on problems in approximation theory. *Mat. Zbirnik KDU*, pages 169–183, 1948. (Ukrainian).

[51] S. Zuhovickiĭ. On approximation of real functions in the sense of P.L. Čebyšev. *AMS Translations of Mathematical Monographs.*, 19(2):221–252, 1962.