

An Epigraphical Approach to the Representer Theorem

Vincent Duval

*INRIA & CEREMADE, Université Paris-Dauphine, Paris, France
vincent.duval@inria.fr*

Received: January 7, 2020

Accepted: June 2, 2020

Describing the solutions of inverse problems arising in signal or image processing is an important issue both for theoretical and numerical purposes. We propose a principle which describes the solutions to convex variational problems involving a finite number of measurements. We discuss its optimality on various problems concerning the recovery of Radon measures.

Keywords: Representer theorem, convex optimization, Carathéodory theorem.

2010 Mathematics Subject Classification: 52A41, 90C25, 90C49.

The recovery of an unknown vector from a finite number of linear measurements is an inverse problem which frequently arises in signal processing or machine learning. Convex variational approaches provide a flexible framework to address this task, typically by solving a problem of the form

$$(\mathcal{P}) \quad \min_{u \in V} R(u) + f(\Phi u),$$

where V is a real vector space (e.g. the space of signals), $\Phi: V \rightarrow \mathbb{R}^M$ is linear, and $R: V \rightarrow \mathbb{R} \cup \{+\infty\}$ and $f: \mathbb{R}^M \rightarrow \mathbb{R} \cup \{+\infty\}$ are two convex functions.

The solutions to $(\mathcal{P})$ are in general not unique, and it is of crucial importance to understand the structure of the solution set $\mathcal{S}$. As noted in [8], the regularization term R determines the building blocks, or *atoms*, which constitute, by convex combinations, the (or some) elements of $\mathcal{S}$. Those atoms are the extreme points of the level sets of R . Besides its theoretical interest, this description makes a strong case for greedy optimization algorithms such as the conditional gradient (a.k.a. Frank-Wolfe) [5, 2], especially in the case where V is infinite-dimensional: even though it is not possible to encode a generic vector of V on a computer, the *atoms* induced by R might be simple enough to be handled (possibly with some approximation) numerically.

The goal of the present note is to state this representation principle as precisely as possible, that is, the description of solutions of $(\mathcal{P})$ as a convex combination of *atoms*, while emphasizing the geometric essence of the property.

1. Related works and main theorem

1.1. Representer theorem

While representation results describing the sparsity (resp. rank) of some solutions of linear (resp. semi-definite) programs are well-known (see e.g. [1, 21]), our main focus is on infinite dimensional problems, for instance the *Basis Pursuit for Measures* [6, 10],

$$\min_{m \in \mathcal{M}(\Omega)} |m|(\Omega) \quad \text{s.t.} \quad \Phi m = y \quad (1)$$

where $y \in \mathbb{R}^M$, Ω is a compact subset of $\mathbb{R}^d$ or the d -dimensional torus $\mathbb{T}^d$, $\mathcal{M}(\Omega)$ is the space of Radon measures on Ω and $|m|(\Omega)$ denotes the total variation of m . The observation is $\Phi m = \left(\int_{\Omega} \varphi_i(x) dm(x) \right)_{1 \leq i \leq M}$ where $\{\varphi_i\}_{1 \leq i \leq M}$ is a family of continuous functions on Ω . It is known [27] that there is a solution which can be represented as

$$m = \sum_{i=1}^r a_i \delta_{x_i} \quad \text{where } r \leq M, \{(x_i, a_i)\}_{i=1}^r \subseteq \Omega \times \mathbb{R}. \quad (2)$$

More precisely, Fisher and Jerome [13] (see also [26, 14]) have proved that *the extreme points of the solution set $\mathcal{S}$ are of the form (2)*. Since, in a locally convex Hausdorff space, each nonempty compact convex set is the closed convex hull of its extreme points, knowing the solutions of the form (2) allows to recover the whole set of solutions.

A related result was proved by L.C. Dubins [11] and V. Klee [18] in the study of convex sets: the extreme points of the intersection of a linearly closed and bounded convex set C with an affine space of codimension M are a convex combination of at most $M + 1$ extreme points of C . In a variational problem with an equality constraint such as (1) (i.e. $f(w) = \chi_{\{y\}}(w) \stackrel{\text{def.}}{=} 0$ if $w = y$, $+\infty$ otherwise), one may regard the solution set as the intersection of the level set $\{R \leq \min \mathcal{P}\}$ with $\Phi^{(-1)}\{y\}$. However, compared to (2), applying the Dubins-Klee theorem yields one atom too many, and in [3] (together with coauthors) we have provided an argument which reduces that number to M by exploiting the structure of convex optimization problems (see also [4]).

The present note describes an alternative argument which handles the case of general convex fidelity terms much more precisely than in [3] and makes the statement more symmetric. Moreover, we discuss the optimality of the theorem in the particular case of Radon measure recovery.

1.2. Main theorem

Our main result relates the dimensions of the faces in the level sets of the regularizer and the fidelity term,

$$\begin{aligned} \{R \leq R(p)\} &\stackrel{\text{def.}}{=} \{u \in V \mid R(u) \leq R(p)\}, \\ \{f \leq f(\Phi p)\} &\stackrel{\text{def.}}{=} \{w \in \mathbb{R}^M \mid f(w) \leq f(\Phi p)\}. \end{aligned}$$

The notion of face used below is described in Section 2.1. Note that, possibly changing f and M , there is no loss of generality in assuming that Φ is surjective.

Theorem 1.1. *Let $R:V \rightarrow \mathbb{R} \cup \{+\infty\}$, $f:\mathbb{R}^M \rightarrow \mathbb{R} \cup \{+\infty\}$ be two convex functions, and let $\Phi:V \rightarrow \mathbb{R}^M$ be linear surjective. Assume that $p \in \mathcal{S}$, $R(p) + f(\Phi p) < +\infty$, and that $\{R \leq R(p)\}$ is linearly closed and contains no line.*

If p belongs to a face of $\mathcal{S}$ with dimension $j < +\infty$, then it belongs to a face of $\{R \leq R(p)\}$ with dimension at most k , where

$$k \stackrel{\text{def.}}{=} \begin{cases} M - \ell + j - 1 & \text{if } (R(p) > \inf R) \text{ or } (f(\Phi p) > \inf f), \\ M - \ell + j & \text{otherwise.} \end{cases} \quad (3)$$

and ℓ is the dimension of the minimal face of Φp in $\{f \leq f(\Phi p)\}$. If, moreover, p satisfies the double obliqueness condition described in Definition 2.2, the number k can be reduced to $M - \ell + j - 2$.

We deduce the following representation for p .

Corollary 1.2. *Under the assumptions of Theorem 1.1, p can be written as a convex combination of (at most) $k + 1$ extreme points of $\{R \leq R(p)\}$, or k points of $\{R \leq R(p)\}$, each an extreme point or a point in an extreme ray.*

Remark 1.3. In the case of an equality constraint ($f = \chi_{\{y\}}$) or a strictly convex function f , we have $\ell = 0$. Hence, p belongs to a face of $\{R \leq R(p)\}$ with dimension (at most) $M + j - 1$ if it belongs to a face of $\mathcal{S}$ with dimension j and $R(p) < \inf R$. This is coherent with [3]. On the other hand, with polyhedral fidelity terms f such as the ℓ^1 or ℓ^∞ norm, non-trivial values of ℓ should be taken into account. $\square$

Remark 1.4. Furthermore, for $j = 0$, i.e. p extreme point of $\mathcal{S}$, we obtain that p belongs to a face of dimension (at most) $M - 1$ provided $R(p) > \inf R$. Hence Corollary 1.2 recovers the Fisher-Jerome theorem. $\square$

2. Proof of the main result

2.1. Reminder on the faces of convex sets.

We first recall a few definitions and basic properties. Given two points x and y in V , we define as usual the closed interval (or line segment) joining x to y as $[x, y] \stackrel{\text{def.}}{=} \{tx + (1 - t)y \mid 0 \leq t \leq 1\}$, and consequently the open interval joining x to y as $]x, y[\stackrel{\text{def.}}{=} [x, y] \setminus \{x, y\}$. A line (resp. an open half line) is a set of the form $a + \mathbb{R}v$ (resp. $\{a + tv \mid t > 0\}$) where $a, v \in V$ and $v \neq 0$. In the following, $C \subseteq V$ denotes a *convex set*, i.e. for any $x, y \in C$, the segment $[x, y]$ lies in C .

The set C is *linearly closed* (resp. linearly bounded) if its intersection with any line is a closed (resp. bounded) subset of that line.

We say that a point $u \in C$ belongs to the *relative algebraic interior* (or *intrinsic core*) of C if

$$\forall v \in \text{Aff } C, \quad \exists \epsilon > 0, \forall \lambda \in]-\epsilon, \epsilon[, \quad u + \lambda(v - u) \in C, \quad (4)$$

where $\text{Aff } C$ denotes the affine hull of C . Equivalently (see [17, Prop. 2.3]), u is in the relative algebraic interior of C if and only if

$$\forall v \in C \setminus \{u\}, \exists z \in C, \quad u \in]v, z[. \quad (5)$$

We say that C is *internal* if it is equal to its relative algebraic interior.

A point $x \in C$ is an *extreme point* of C if there is no open interval in C containing x , or equivalently if $C \setminus \{x\}$ is convex. An *extreme ray* ρ of C is a half-line contained in C such that any open interval I which intersects ρ must satisfy $I \subseteq \rho$. More generally, a subset F of C is said to be a *face* of C if F is convex and for all open interval $I \subseteq C$ which intersects F , $I \subset F$. An alternative definition of an extreme point is “a point x such that $\{x\}$ is a face of C ”. Similarly, extreme rays may be defined as the half-lines which are a face of C . The dimension of a face, $\dim F$, is defined as the dimension of its affine hull $\text{Aff}(F)$.

A canonical choice of face is given by the notion of *elementary face*. Given $x \in C$, we define $\mathcal{F}(x, C)$ as the intersection of all the faces of C which contain x . It is also a face, hence it is the *minimal face* of C (for the inclusion) which contains x . We call such sets the elementary faces of C . It turns out that $\mathcal{F}(x, C)$ is equal to the largest internal subset of C which contains x (see [11, Th. 2.1]), hence it is the union of $\{x\}$ and all the open intervals of C which contain x . Moreover, the elementary faces yield a *partition* of C with $y \in \mathcal{F}(x, C)$ if and only if $\mathcal{F}(x, C) = \mathcal{F}(y, C)$.

The behavior of the elementary faces when performing several operations on convex sets is described below.

Intersection. Since the elementary face $\mathcal{F}(x, C)$ is the union of $\{x\}$ and all the open intervals of a convex set which contain x , one may check that if C_1 and C_2 are two convex sets,

$$\mathcal{F}(x, C_1 \cap C_2) = \mathcal{F}(x, C_1) \cap \mathcal{F}(x, C_2). \quad (6)$$

Moreover, if $W_{1,2}$, W_1 , W_2 respectively denote the affine hulls of those faces, they consist in the collection of lines through x which respectively intersect $C_1 \cap C_2$, C_1 , C_2 through an open interval. As a consequence,

$$W_{1,2} = W_1 \cap W_2. \quad (7)$$

Cartesian product. If C_1 , C_2 are convex subsets of the vector spaces V_1 , V_2 , it is possible to check that $\mathcal{F}(x_1, C_1) \times \mathcal{F}(x_2, C_2)$ is both a face of $C_1 \times C_2$ and an internal set. As a result,

$$\mathcal{F}((x_1, x_2), C_1 \times C_2) = \mathcal{F}(x_1, C_1) \times \mathcal{F}(x_2, C_2). \quad (8)$$

Moreover, if $W_{1,2}$, W_1 , W_2 respectively denote the affine hulls of the above-mentioned faces, it holds

$$W_{1,2} = W_1 \times W_2. \quad (9)$$

Affine map. If $\psi: V_1 \rightarrow V_2$ is an affine bijective map, it preserves the elementary faces:

$$\mathcal{F}(\psi(x), \psi(C)) = \psi(\mathcal{F}(x, C)). \quad (10)$$

If W_1 (resp. W_2) denotes the affine hull $\mathcal{F}(x, C)$ (resp. $\mathcal{F}(\psi(x), \psi(C))$),

$$W_2 = \psi(W_1). \quad (11)$$

2.2. Epigraphical reformulation

Now, we proceed with the proof of Theorem 1.1. We consider p as in Theorem 1.1, and we write $f_\Phi \stackrel{\text{def.}}{=} f \circ \Phi$, $t^\star \stackrel{\text{def.}}{=} R(p) + f_\Phi(p)$. Instead of directly studying $\mathcal{S}$ and the

level sets $\{R \leq R(p)\}$, $\{f_\Phi \leq f_\Phi(p)\}$, we consider $\hat{\mathcal{S}} \stackrel{\text{def.}}{=} \{(u, R(u)) \mid u \in \mathcal{S}\}$, as well as the epigraph of R and the hypograph of $t^* - f_\Phi$,

$$\hat{\mathcal{E}} \stackrel{\text{def.}}{=} \text{epi}(R) \stackrel{\text{def.}}{=} \{(u, r) \in V \times \mathbb{R} \mid R(u) \leq r\},$$

$$\hat{\mathcal{H}} \stackrel{\text{def.}}{=} \text{hypo}(t^* - f_\Phi) \stackrel{\text{def.}}{=} \{(u, r) \in V \times \mathbb{R} \mid t^* - f_\Phi(u) \geq r\}.$$

We note that $\hat{\mathcal{E}}$ and $\hat{\mathcal{H}}$ are convex, with $\hat{\mathcal{S}} = \hat{\mathcal{E}} \cap \hat{\mathcal{H}}$ (see Figure 2.1 for an illustration).

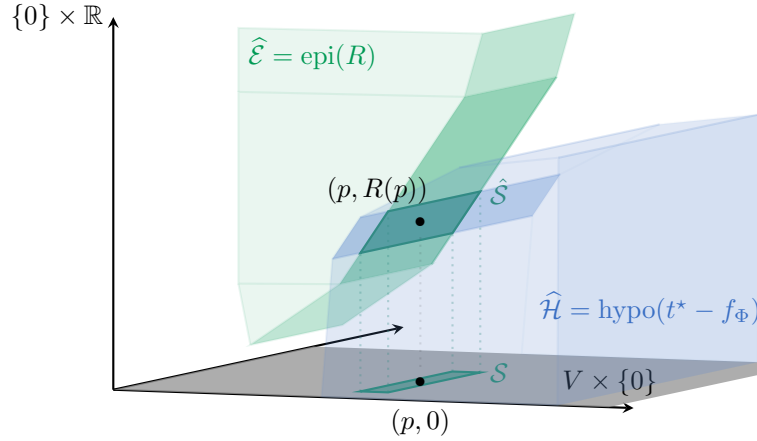


Figure 2.1: The solution set $\mathcal{S}$ is equivalent, up to an affine isomorphism, to the set $\hat{\mathcal{S}}$ (see Section 2.1).

Indeed, if $(u, r) \in \hat{\mathcal{E}} \cap \hat{\mathcal{H}}$, then $R(u) \leq r \leq t^* - f_\Phi(u)$, hence

$$R(u) + f_\Phi(u) \leq r + f_\Phi(u) \leq t^*.$$

Since $t^* = \inf(\mathcal{P})$, the left-hand side is bounded below by t^* , hence $u \in \mathcal{S}$ and $r = R(u)$. This proves that $\hat{\mathcal{E}} \cap \hat{\mathcal{H}} \subseteq \hat{\mathcal{S}}$. The converse inclusion is straightforward.

Using (6), we get

$$\mathcal{F}((p, R(p)), \hat{\mathcal{E}} \cap \hat{\mathcal{H}}) = \mathcal{F}((p, R(p)), \hat{\mathcal{E}}) \cap \mathcal{F}((p, R(p)), \hat{\mathcal{H}}). \quad (12)$$

To understand the dimension of those faces, we need to consider their affine hulls. Let $\hat{S}$, $\hat{E}$ and $\hat{H}$ be the respective affine hulls of

$$\mathcal{F}((p, R(p)), \hat{\mathcal{E}} \cap \hat{\mathcal{H}}), \quad \mathcal{F}((p, R(p)), \hat{\mathcal{E}}) \quad \text{and} \quad \mathcal{F}((p, R(p)), \hat{\mathcal{H}}).$$

From (7), we have $\hat{S} = \hat{E} \cap \hat{H}$. Up to a translation of the origin in $V \times \mathbb{R}$, we assume without loss of generality that $(p, R(p)) = (0, 0)$, so that *all the above-mentioned affine hulls are now linear hulls*. By classical results of linear algebra, we get $\dim \hat{E} = \dim(\hat{E} \cap \hat{H}) + \text{codim}_{\hat{E}}(\hat{E} \cap \hat{H})$ and

$$\text{codim}_{\hat{E}}(\hat{E} \cap \hat{H}) = \text{codim}_{\hat{E} + \hat{H}}(\hat{H}) = \text{codim}_{V \times \mathbb{R}} \hat{H} - \text{codim}_{V \times \mathbb{R}}(\hat{E} + \hat{H}).$$

Combining those equalities, we get

$$\dim \hat{E} = \text{codim}_{V \times \mathbb{R}} \hat{H} + \dim(\hat{S}) - \text{codim}_{V \times \mathbb{R}}(\hat{E} + \hat{H}). \quad (13)$$

2.3. Faces in epigraph and in level sets

To relate the faces of $\hat{\mathcal{S}}$ and $\mathcal{S}$, we note that $\hat{\mathcal{S}}$ is the image of $\mathcal{S}$ by some injective linear map.

Lemma 2.1. *There is a linear map $L : \text{Span}(\mathcal{S}) \rightarrow \mathbb{R}$ such that R coincides with L on $\mathcal{S}$. Moreover, the map $\hat{L} : \text{Span}(\mathcal{S}) \rightarrow \text{Span}(\hat{\mathcal{S}})$ defined by $\hat{L}(v) \stackrel{\text{def.}}{=} (v, L(v))$ is bijective and $\hat{L}(\mathcal{S}) = \hat{\mathcal{S}}$.*

The proof consists in observing that the convex function R must coincide with the concave function $t^* - f_\Phi$ on $\mathcal{S}$. We omit it for brevity. As a consequence of Lemma 2.1, we obtain $j = \dim \mathcal{F}(0, \mathcal{S}) = \dim \mathcal{F}((0, 0), \hat{\mathcal{S}}) = \dim(\hat{\mathcal{S}})$.

As for $\hat{\mathcal{E}}$ and $\{R \leq 0\}$, we have $\hat{\mathcal{E}} \cap (V \times \{0\}) = \{R \leq 0\} \times \{0\}$, and using (6) and (8) we obtain

$$\mathcal{F}((0, 0), \hat{\mathcal{E}}) \cap (V \times \{0\}) = \mathcal{F}((0, 0), \hat{\mathcal{E}} \cap (V \times \{0\})) = \mathcal{F}(0, \{R \leq 0\}) \times \{0\}. \quad (14)$$

Let $E \stackrel{\text{def.}}{=} \text{Span}(\mathcal{F}(0, \{R \leq 0\})) \subseteq V$. From (7), we note that the linear spans $\hat{E}$ and E are related through $\hat{E} \cap (V \times \{0\}) = E \times \{0\}$. As a result

$$\dim \hat{E} = \dim E + \text{codim}_{\hat{E}}(E \times \{0\}), \quad (15)$$

where $\text{codim}_{\hat{E}}(E \times \{0\}) \in \{0, 1\}$. If $\text{codim}_{\hat{E}}(E \times \{0\}) = 0$, we say that the face $\mathcal{F}((0, 0), \hat{\mathcal{E}})$ is *horizontal*. Otherwise we say that it is *oblique*.

Now, we examine $\hat{\mathcal{H}}$ and $\{f \leq f(0)\}$. Let Z be a linear complement to $\ker \Phi$ in V . Since $\text{rank } \Phi = M$, the restriction $\Phi|_Z : Z \rightarrow \mathbb{R}^M$ is an isomorphism. As a result, the mapping

$$\begin{aligned} \psi : \quad V \times \mathbb{R} &\longrightarrow \ker \Phi \times Z \times \mathbb{R} \longrightarrow (\ker \Phi) \times (\mathbb{R}^M \times \mathbb{R}) \\ (u, r) &\longmapsto (k, z, r) \longmapsto (k, (\Phi z, r)) \end{aligned}$$

(where (k, z) is the unique element in $\ker \Phi \times Z$ such that $u = k + z$) is an isomorphism. In particular, since ψ maps $\hat{\mathcal{H}}$ to $\ker \Phi \times (\text{hypo}(t^* - f))$,

$$\begin{aligned} \psi \left(\mathcal{F}((u, r), \hat{\mathcal{H}}) \right) &= \mathcal{F}(\psi(u, r), \psi(\hat{\mathcal{H}})) \\ &= \mathcal{F}((k, (\Phi z, r)), \ker \Phi \times (\text{hypo}(t^* - f))) \\ &= \ker \Phi \times \mathcal{F}((\Phi z, r), \text{hypo}(t^* - f)). \end{aligned}$$

Applying this to $(u, r) = (p, R(p)) = (0, 0)$ and considering the linear spans, we obtain

$$\text{codim}_{V \times \mathbb{R}} \hat{H} = M + 1 - \hat{\ell}, \quad (16)$$

where $\hat{\ell}$ is the dimension of $\mathcal{F}((0, 0), \text{hypo}(t^* - f))$. Just like the faces of the regularizer, we may define $H = \text{Span } \mathcal{F}(0, \{f \leq f(0)\})$ and notice that

$$\dim \hat{H} = \dim H + \text{codim}_{\hat{H}}(H \times \{0\}).$$

In consequence we have $\hat{\ell} = \dim \mathcal{F}(0, \{f \leq f(0)\})$ (in this case we say that the face $\mathcal{F}((0, 0), \text{hypo}(t^* - f))$ is *horizontal*), or $\hat{\ell} = \dim \mathcal{F}(0, \{f \leq f(0)\}) + 1$ (in this case we say it is *oblique*).

From the above discussion, we see that the horizontality or obliqueness of the faces in $\widehat{\mathcal{E}}$ and $\widehat{\mathcal{H}}$ play an important role. The following condition will be useful.

Definition 2.2. We say that p satisfies the *double obliqueness condition* if both $\mathcal{F}((p, R(p)), \widehat{\mathcal{E}})$ and $\mathcal{F}((\Phi p, t^* - f(\Phi p)), \text{hypo}(t^* - f))$ are oblique. In other words,

$$\text{codim}_{\widehat{E}}(E \times \{R(p)\}) = 1 \quad \text{and} \quad \text{codim}_{\widehat{H}}(H \times \{R(p)\}) = 1. \quad (17)$$

2.4. Conclusion of the proof of Theorem 1.1

It remains to study the last term of (13).

Lemma 2.3. *The following inequality holds: $\text{codim}_V(\widehat{E} + \widehat{H}) \geq 1$.*

If both $\mathcal{F}((0, 0), \widehat{\mathcal{E}})$ and $\mathcal{F}((0, 0), \text{hypo}(t^ - f))$ are horizontal and $R(0) > \inf R$ or $f_\Phi(0) > \inf f_\Phi$, then $\text{codim}_{V \times \mathbb{R}}(\widehat{E} + \widehat{H}) \geq 2$.*

Proof. First, in the general case, we prove that $\text{codim}_V(\widehat{E} + \widehat{H}) \geq 1$ by arguing that $(\{0\} \times \mathbb{R}) \cap (\widehat{E} + \widehat{H}) = \{0\}$. Indeed, assume by contradiction that $(\{0\} \times \mathbb{R}) \subseteq (\widehat{E} + \widehat{H})$. Then, for all $\epsilon > 0$, there exists a vector $(e, \alpha) \in \widehat{E}$, a vector $(h, \beta) \in \widehat{H}$, such that $(0, -\epsilon) = (e, \alpha) + (h, \beta)$. Possibly reducing ϵ , we may assume that (e, α) and (h, β) are so small that $(e, \alpha) \in \mathcal{F}((0, 0), \widehat{\mathcal{E}})$ and $(h, \beta) \in \mathcal{F}((0, 0), \widehat{\mathcal{H}})$. Arguing as in Section 2.1, it is possible to prove that R (resp. $t^* - f_\Phi$) coincides with a linear function on $\{e' \mid (e', \alpha') \in \mathcal{F}((0, 0), \widehat{\mathcal{E}})\}$ (resp. on $\{h' \mid (h', \beta') \in \mathcal{F}((0, 0), \widehat{\mathcal{H}})\}$), with $R(e') = \alpha'$ and $\beta' = t^* - f_\Phi(h')$.

$$\begin{aligned} \text{Hence,} \quad (R + f_\Phi)(e) &= R(e) + f_\Phi(-h) = R(e) + (f_\Phi(-h) - t^*) + t^* \\ &= \alpha + \beta + t^* = t^* - \epsilon < t^*, \end{aligned}$$

which contradicts the fact that t^* is the minimal value of $(\mathcal{P})$.

Now, we assume that both $\mathcal{F}((0, 0), \widehat{\mathcal{E}})$ and $\mathcal{F}((0, 0), \widehat{\mathcal{H}})$ are horizontal and that $R(0) > \inf R$ (the case $f_\Phi(0) > \inf f_\Phi$ being similar); we prove $\text{codim}_{V \times \mathbb{R}}(\widehat{E} + \widehat{H}) \geq 2$.

Note that, both faces being horizontal, we have $\widehat{E} + \widehat{H} \subseteq (V \times \{0\})$. As a result $\widehat{E} = E \times \{0\}$ and $\widehat{H} = H \times \{0\}$, and it is sufficient to prove that $E + H$ has a nontrivial complement in V . By assumption, there exists $x' \in V$ such that $R(x') < R(0)$. By contradiction, if $E + H = V$, we may write $x' = e + h$ with $e \in E$, $h \in H$. For $\theta > 0$, we consider the point

$$x_\theta \stackrel{\text{def.}}{=} \frac{\theta}{1 + \theta} x' + \frac{1}{1 + \theta} (-\theta e) = \frac{\theta}{1 + \theta} h.$$

Since 0 is internal to $\mathcal{F}(0, \{R \leq 0\})$ (resp. $\mathcal{F}(0, \{f_\Phi \leq f_\Phi(0)\})$), for $\theta > 0$ small enough, $(-\theta e) \in \mathcal{F}(0, \{R \leq 0\})$ (resp. $\frac{\theta}{1 + \theta} h \in \mathcal{F}(0, \{f_\Phi \leq f_\Phi(0)\})$). Hence,

$$(R + f_\Phi)(x_\theta) \leq \frac{\theta}{1 + \theta} \underbrace{R(x')}_{< R(0)} + \frac{1}{1 + \theta} \underbrace{R(-\theta e)}_{= R(0)} + f_\Phi(x_\theta) < (R + f_\Phi)(0), \quad (18)$$

which contradicts the fact that 0 is a minimizer.

As a result, $E + H \subsetneq V$. If W is a linear complement to $E + H$ in V , then $(W \times \{0\}) \oplus (\{0\} \times \mathbb{R})$ is thus a linear complement to $\widehat{E} + \widehat{H}$ in $V \times \mathbb{R}$, with dimension ≥ 2 . $\square$

Now, we can finish the proof of Theorem 1.1. From (13) and (16), we note that

$$\dim \widehat{E} = M + 1 - \hat{\ell} + j - \text{codim}_V(\widehat{E} + \widehat{H}).$$

First, we assume that both $\mathcal{F}((0, 0), \widehat{\mathcal{E}})$ and $\mathcal{F}((0, 0), \text{hypo}(t^* - f))$ are horizontal: $\dim \widehat{E} = \dim E$ and $\hat{\ell} = \ell$. If $R(0) = \inf R$ and $f_\Phi(0) = \inf f_\Phi$, the inequality of Lemma 2.3 yields $\dim E \leq M - \ell + j$. On the other hand, if $R(0) > \inf R$ or $f_\Phi(0) > \inf f_\Phi$, the stronger inequality of Lemma 2.3 yields $\dim E \leq M - \ell + j - 1$. Now, if $\mathcal{F}((0, 0), \widehat{\mathcal{E}})$ is oblique, then $\dim \widehat{E} = \dim E + 1$ and $R(0) > \inf R$.

If $\mathcal{F}((0, 0), \text{hypo}(t^* - f))$ is oblique, then $\hat{\ell} = \ell + 1$ and $f(0) > \inf f$. Using that $\text{codim}_V(\widehat{E} + \widehat{H}) \geq 1$, we obtain $\dim E \leq M - \ell + j - 1$ if only one of the faces is oblique, and $\dim E \leq M - \ell + j - 2$ if both are oblique. This concludes the proof.

2.5. Proof of Corollary 1.2

We derive the representation of $p = 0$ as a convex combination. Let $k < +\infty$ be the above-mentioned upper-bound on $\dim E$ and let F be the linear closure of $\mathcal{F}(0, \{R \leq R(0)\})$. As $\mathcal{F}(0, \{R \leq R(0)\}) \subseteq E$, we know that F is in fact its closure for the finite-dimensional topology of E . By the Carathéodory-Klee theorem [18, Th. 3], since F is closed and contains no line, its points are convex combinations of at most $k + 1$ extreme points of F or k points, each an extreme point or a point in an extreme ray. To conclude, we note that F is a face of $\{R \leq R(0)\}$, its extreme points are thus extreme points of $\{R \leq R(0)\}$. $\square$

3. Level sets containing lines

3.1. The generalized splines problem.

In [13], Fisher and Jerome have in fact considered the more general problem

$$\min_{u \in V} |Lu|(\Omega) \quad \text{s.t.} \quad Lu \in \mathcal{M}(\Omega) \quad \text{and} \quad \Phi u = y, \quad (19)$$

where $V \subseteq \mathcal{D}'(\Omega)$ is a suitably defined space of distributions, and the differential operator $L : \mathcal{D}'(\Omega) \rightarrow \mathcal{D}'(\Omega)$ maps V onto $\mathcal{M}(\Omega)$. We refer to [13] for precise assumptions and to [26, 14, 15] for extensions to a more general setting. Under some topological assumptions, the results in [26, 14] describe the extreme points of the solution set to (19) as generalized splines

$$u(x) = \sum_{k=1}^r a_k L^+ \delta_{x_k} + P(x) \quad (20)$$

where $a_k \in \mathbb{R}$, L^+ is some pseudo-inverse of L (one may see $L^+ \delta_{x_k}$ as a Green function for L) and $P \in \text{Ker } L$. Moreover, $r \leq M - \dim \Phi(\text{ker } L)$.

For $\Omega = \mathbb{R}$ and $L = D^n$, where D denotes the differentiation operator, one obtains the polynomial splines

$$u(x) = \sum_{k=1}^r \frac{(x - x_k)_+^n}{n!} + P(x),$$

where $P \in \mathbb{R}_{n-1}[X]$. In this section, we examine how our discussion can be extended to the case of level set containing lines so as to obtain representations of the form (20).

3.2. Convex sets and their lineality space

We first need to recall several properties of convex sets containing lines (see for instance [17] or [24, Ch.8]). We say that a nonempty convex set $C \subseteq V$ is invariant in the direction $v \in V$ if

$$C + \mathbb{R}v \subseteq C. \quad (21)$$

The collection of all vectors $v \in V$ such that (21) holds is a vector space called the *lineality space* of C , denoted by $\text{lin } C$.

If C is internal or linearly closed, given $v \in V \setminus \{0\}$, it is equivalent to say that C is invariant in the direction v , or to say that C contains a line directed by v , i.e. $(x_0 + \mathbb{R}v) \subseteq C$ for some $x_0 \in V$. As a consequence, if C_1, C_2 are two nonempty convex sets, then $\text{lin } C_1 \cap \text{lin } C_2 \subseteq \text{lin } C_1 \cap C_2$, with equality if C_1 and C_2 are internal or linearly closed.

If C is invariant by some subspace K , i.e. $K \subseteq \text{lin } C$, it is sometimes convenient to quotient the ambient space by K . Considering a linear complement¹ W to K , we note that there is a linear isomorphism

$$\begin{aligned} \psi : \quad V &\longrightarrow K \times W \longrightarrow K \times (V/K) \\ u &\longmapsto (k, w) \longmapsto (k, \pi_K(w)) \end{aligned}$$

where (k, w) is the unique element in $K \times W$ such that $u = k + w$, and that $\pi_K : V \rightarrow V/K$ is the canonical surjection. In other words $\pi_K(w) = \pi_K(u) \stackrel{\text{def.}}{=} u + K$ is the coset of u .

From (10), we note that $\psi(\mathcal{F}(u, C)) = K \times \mathcal{F}(\pi_K(u), \pi_K(C))$ so that the faces of $\pi_K(C)$ can be easily deduced from those of C and conversely, with

$$\pi_K(\mathcal{F}(u, C)) = \mathcal{F}(\pi_K(u), \pi_K(C)). \quad (22)$$

If C is internal (resp. linearly closed), then $\pi_K(C)$ is internal (resp. linearly closed) and, if $K = \text{lin } C$, $\pi_K(C)$ contains no line.

3.3. Back to the optimization problem

Let $K = \text{lin } \{R \leq R(p)\}$ and $N \stackrel{\text{def.}}{=} \ker \Phi$.

We note that $\mathcal{F}(p, \mathcal{S})$ is invariant by $K \cap N$.

¹We use freely the axiom of choice, hence any subspace of V admits a complement subspace.

Indeed, the face of the epigraph $\mathcal{F}((p, R(p)), \hat{\mathcal{E}})$ is internal and contains the set $K \times \{R(p)\}$, hence it is invariant by $\hat{K} \stackrel{\text{def.}}{=} K \times \{0\}$. On the other hand, the hypograph $\hat{\mathcal{H}}$ (hence $\mathcal{F}((p, R(p)), \hat{\mathcal{H}})$) is invariant by $\hat{N} \stackrel{\text{def.}}{=} N \times \{0\}$. From (12) we deduce that $\mathcal{F}((p, R(p)), \hat{\mathcal{E}} \cap \hat{\mathcal{H}})$ is invariant by $\hat{K} \cap \hat{N} = (K \cap N) \times \{0\}$. Since $\mathcal{F}(p, \mathcal{S})$ is its projection onto the horizontal hyperplane (see Section 2.1), it is invariant by $K \cap N$.

Therefore, $\mathcal{F}(p, \mathcal{S})$ is linearly isomorphic to $(K \cap N) \times \pi_{K \cap N}(\mathcal{F}(p, \mathcal{S}))$. Instead of considering the dimension of $\mathcal{F}(p, \mathcal{S})$ to describe the point p , the following theorem relies on the dimension of the coset $\pi_{K \cap N}(\mathcal{F}(p, \mathcal{S}))$.

Theorem 3.1. *Let $R : V \rightarrow \mathbb{R} \cup \{+\infty\}$, $f : \mathbb{R}^M \rightarrow \mathbb{R} \cup \{+\infty\}$ be two convex functions, and let $\Phi : V \rightarrow \mathbb{R}^M$ be linear surjective. Assume that $p \in \mathcal{S}$, such that $R(p) + f(\Phi p) < +\infty$, and that $\{R \leq R(p)\}$ is linearly closed. Finally, let $K \stackrel{\text{def.}}{=} \text{lin} \{R \leq R(p)\}$, $d \stackrel{\text{def.}}{=} \dim \Phi(K)$, and $N \stackrel{\text{def.}}{=} \ker \Phi$.*

If $\dim(\pi_{K \cap N}(\mathcal{F}(p, \mathcal{S}))) = j < +\infty$, then $\pi_K(p)$ belongs to a face of $\pi_K(\{R \leq R(p)\})$ with dimension at most k , where

$$k \stackrel{\text{def.}}{=} \begin{cases} M - \ell + j - d - 1 & \text{if } (R(p) > \inf R) \text{ or } (f(\Phi p) > \inf f), \\ M - \ell + j - d & \text{otherwise,} \end{cases} \quad (23)$$

and ℓ is the dimension of the minimal face of Φp in $\{f \leq f(\Phi p)\}$.

In particular, $\pi_K(p)$ can be written as a convex combination of (at most):

- $k + 1$ extreme points of $\pi_K(\{R \leq R(p)\})$,
- or k points of $\pi_K(\{R \leq R(p)\})$, each an extreme point or a point in an extreme ray.

If, moreover, p satisfies the obliqueness condition described in Definition 2.2, the number k can be reduced to $M - \ell + j - d - 2$.

In particular, if $p_1, \dots, p_r \in \{R \leq R(p)\}$ are such that $\pi_K(p_1), \dots, \pi_K(p_r)$ denote those extreme points (or points in extreme rays),

$$p = \sum_{i=1}^r \theta_i p_i + u_K, \quad \text{where } \theta_i \geq 0, \sum_{i=1}^r \theta_i = 1, \quad \text{and } u_K \in K. \quad (24)$$

Remark 3.2. In practice, if $\hat{\mathcal{E}}$ is linearly closed (e.g. if R is lower semi-continuous for some vector topology), then the whole solution set $\mathcal{S}$ is invariant by $(K \cap N)$, and $\pi_{K \cap N}(\mathcal{F}(p, \mathcal{S})) = (\mathcal{F}(\pi_{K \cap N}(p), \pi_{K \cap N}(\mathcal{S})))$. However, notice that the solution set $\mathcal{S}$ may have more invariant directions than just $(K \cap N)$. $\square$

Proof. We assume, up to a change of the origin in $V \times \mathbb{R}$, that $(p, R(p)) = (0, 0)$.

As noted above, $\mathcal{F}((p, R(p)), \hat{\mathcal{E}})$ is invariant by $\hat{K} = K \times \{0\}$, hence $\hat{K} \subseteq \hat{E}$.

Similarly, $\hat{N} \subseteq \hat{H}$, and classical results on quotient spaces [20, Ch. 3, Sec. 1] imply

$$\left\{ \begin{array}{l} \hat{E}/\hat{K} \approx \left(\hat{E}/(\hat{K} \cap \hat{N}) \right) / \left(\hat{K}/(\hat{K} \cap \hat{N}) \right) \\ \hat{E}/(\hat{E} \cap \hat{H}) \approx \left(\hat{E}/(\hat{K} \cap \hat{N}) \right) / \left((\hat{E} \cap \hat{H})/(\hat{K} \cap \hat{N}) \right) \end{array} \right. \quad (25)$$

hence, provided the corresponding dimensions are finite (what we shall prove below),

$$\begin{aligned} & \dim(\hat{E}/\hat{K}) \\ &= \dim(\hat{E}/(\hat{E} \cap \hat{H})) + \dim((\hat{E} \cap \hat{H})/(\hat{K} \cap \hat{N})) - \dim(\hat{K}/(\hat{K} \cap \hat{N})). \end{aligned} \quad (26)$$

The term $\dim(\hat{E}/(\hat{E} \cap \hat{H})) = \text{codim}_{\hat{E}}(\hat{E} \cap \hat{H})$ has already been studied in Sections 2.2 and 2.3. Now, we prove that $\dim((\hat{E} \cap \hat{H})/(\hat{K} \cap \hat{N})) = j$. Since $\hat{L}$ (see Lemma 2.1) satisfies $\hat{L}(K \cap N) = (K \cap N) \times \{0\}$, it induces a map $\hat{\hat{L}}$ which makes the following commutative diagram.

$$\begin{array}{ccc} S & \xrightarrow{\hat{L}} & \hat{S} \\ \downarrow \pi_{K \cap N} & & \downarrow \pi_{\hat{K} \cap \hat{N}} \\ S/K \cap N & \xrightarrow{\hat{\hat{L}}} & \hat{S}/\hat{K} \cap \hat{N} \end{array}$$

Since $\hat{L}$ is bijective, $\hat{\hat{L}}$ is surjective by construction. Let us observe that it is also injective.

Consider any coset in $S/K \cap N$, say $\pi_{K \cap N}(s)$ for some $s \in S$. If $\hat{\hat{L}}(\pi_{K \cap N}(s)) = 0$, it means that $\pi_{\hat{K} \cap \hat{N}}(\hat{L}(s)) = 0$, that is $(s, L(s)) \in (K \times N) \times \{0\}$. Hence $\pi_{K \cap N}(s) = 0$, and $\hat{\hat{L}}$ is injective.

As a result $\dim(\hat{S}/\hat{K} \cap \hat{N}) = \dim(S/K \cap N)$. But by linearity,

$$S/K \cap N = \pi_{K \cap N}(S) = \pi_{K \cap N}(\text{Span}(\mathcal{F}(0, \mathcal{S}))) = \text{Span}(\pi_{K \cap N}(\mathcal{F}(0, \mathcal{S}))),$$

so that $\dim(\hat{S}/\hat{K} \cap \hat{N}) = j$.

It remains to identify $\dim(\hat{K}/(\hat{K} \cap \hat{N}))$. By the first isomorphism theorem, if $\Phi|_K$ denotes the restriction of Φ to K ,

$$\text{Im } \Phi|_K \approx (K/\ker \Phi|_K) = (K/K \cap N) \approx (\hat{K}/\hat{K} \cap \hat{N}).$$

As a result, $\dim(\hat{K}/\hat{K} \cap \hat{N}) = \dim \Phi(K) = d$.

To conclude the proof, we deal with $\hat{E}$. As in Section 2.3, we note that we have $\dim(\hat{E}/\hat{K}) = \dim(E/K) + (0 \text{ or } 1)$, depending on whether $\mathcal{F}(0, \hat{\mathcal{E}})$ is horizontal or oblique.

Hence, $\dim(E/K) \leq k$ where k is given by (23). Moreover,

$$E/K = \pi_K(\text{Span}(\mathcal{F}(0, \{R \leq 0\}))) = \text{Span}(\pi_K(\mathcal{F}(0, \{R \leq 0\})))$$

with the consequence that $\dim(E/K)$ is the dimension of the elementary face of $\pi_K(p)$ in $\pi_K(\{R \leq R(p)\})$. As $\{R \leq R(p)\}$ is linearly closed and K is its lineality space, we note that $\pi_K(\{R \leq R(p)\})$ is linearly closed and contains no line. The representation of $\pi_K(p)$ as a convex combination follows from the same argument as in the proof of Corollary 1.2. $\square$

3.4. Link with the Fisher-Jerome result

The conclusions of Theorem 3.1 and (24) recover the representation result (20) for the Fisher-Jerome problem. Indeed, assuming that there is a point p such that $j = 0$ (e.g. if $\pi_{K \cap N}(\mathcal{S})$ is compact for some locally convex Hausdorff topology), we obtain that $\pi_K(p)$ belongs to a face with dimension at most $M - d - 1$. The quantities $\frac{a_k}{|a_k|} L^+ \delta_{x_k}$ are some specific choices of points p_k such that $\pi_K(p_k) = \frac{a_k}{|a_k|} \delta_{x_k}$ is an extreme point of the total variation unit ball. The operator L^+ is determined by choosing some specific linear complement W to K . Conversely, if W (hence L^+) is fixed and $p_1, \dots, p_r$ are given as in (24), one can recover (20) by adding some elements of K (i.e. changing u_K).

4. Is it optimal?

4.1. Optimality in the general class of convex problems

For each different case of Theorem 1.1 and Corollary 1.2, it is possible to exhibit a convex function for which the upper-bounds are attained. In that sense, the results stated above are optimal. Moreover, due to their geometric essence and the weakness of their topological assumptions², they are quite general. In comparison, an approach relying on the subdifferential and optimality conditions would require a constraint qualification argument together with a suitable choice of topology. Moreover, it is not clear to the author whether exploiting the subdifferential could describe precisely the extreme points that are not exposed.

On the other hand, for several specific problems that we discuss below, the bounds provided by Theorem 1.1 are not tight.

4.2. The Carathéodory number

A first example, discussed in [3], is the case of Semi-Definite Programs (SDP), such as the feasibility problem

$$\min_Q \chi_{S_n^+(\mathbb{R})}(Q) \quad \text{s.t.} \quad \Phi Q = y,$$

²The notions of linearly closed set and algebraic interior used in the present paper can be regarded as topological, since a set is linearly closed (resp. algebraically open) if and only if it is closed (resp. open) in the *core topology* (see [16]). However, in practice, that topology is rarely used since it is not compatible with the vector space structure of V , so that one might as well regard those notions as algebraic definitions. In some cases, a convex set is linearly closed if and only if it is closed for the *convex core topology*, i.e. the finest locally convex topology (see [16, Prop. 4.6 and 8.7]). In any case, linear closedness is a weak assumption, since it is implied by being closed in a topological vector space.

where $\mathcal{S}_n^+(\mathbb{R})$ denotes the cone of symmetric positive semi-definite matrices of size n and, for a given set A , $\chi_A(x) \stackrel{\text{def.}}{=} 0$ if $x \in A$, and $+\infty$ otherwise.

While Corollary 1.2 predicts that an extreme point of the solution set is a convex combination of at most M rank-one matrices (i.e. the points in the extreme rays of $\mathcal{S}_n^+(\mathbb{R})$), it is known [1] that there is a solution which is made of at most $\frac{1}{2}(\sqrt{8M+1}-1) \leq M$ such atoms. As it turns out, even though the solution does belong to a face of $\mathcal{S}_n^+(\mathbb{R})$ with dimension M , the Carathéodory-Klee theorem is not sharp, since *one needs less than M points in extreme rays to represent the points of that face*. That phenomenon is quite common with non-polyhedral sets: for instance, in the Euclidean closed unit ball, every point is a convex combination of at most two extreme points (regardless of the ambient dimension). More generally, if $F \subseteq V$ is convex, linearly bounded, linearly closed and each point of F is internal or an extreme point (e.g. if F is strictly convex), the same property holds.

On the other hand, if $F \subseteq V$ is an M -dimensional convex polytope, it is possible to check that almost every point of F (in the sense of the Lebesgue measure) is a convex combination of $M+1$ (and not less) extreme points of F .

4.3. Nonnegative measures

Another interesting case is the truncated trigonometric moment problem,

$$\min_m \chi_{\mathcal{M}^+(\mathbb{T})}(m) \quad \text{s.t.} \quad \int_{\mathbb{T}} \varphi_k(t) dm(t) = y_k \quad (0 \leq k \leq 2f_c), \quad (27)$$

where $\mathcal{M}^+(\mathbb{T})$ is the set of nonnegative measures on the torus $\mathbb{T} = \mathbb{R}/\mathbb{Z}$, $y \in \mathbb{R}^{2f_c+1}$, and $\varphi_0(t) = 1$, $\varphi_{2j-1}(t) = \cos(j2\pi t)$ and $\varphi_{2j}(t) = \sin(j2\pi t)$ for $1 \leq j \leq f_c$.

The Carathéodory-Toeplitz theorem [7, 25] states that there is a solution to (27) if and only if the matrix

$$T(c) \stackrel{\text{def.}}{=} \begin{pmatrix} c_0 & c_1 & \cdots & c_{f_c} \\ c_1^* & c_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & c_1 \\ c_{f_c}^* & \cdots & c_1^* & c_0 \end{pmatrix}, \quad \text{where} \quad c_j \stackrel{\text{def.}}{=} y_{2j-1} - i y_{2j}, \quad c_0 \stackrel{\text{def.}}{=} y_0,$$

is positive semi-definite. If $r \stackrel{\text{def.}}{=} \text{rank } T(c) \leq f_c$, the solution m is unique, and its support has cardinality r . If $T(c)$ is invertible, there is an infinity of solutions with cardinality $f_c + 1$ (and more). In particular for any $t_0 \in \mathbb{T}$ there is a solution which charges $\{t_0\}$. Note that similar results hold for T-systems on an interval [19, Ch. 4, Sec. 4].

That result contrasts with Corollary 1.2 which would predict a sum of at most $2f_c + 1$ Dirac masses. Here, the situation is different from the case of $\mathcal{S}_n^+(\mathbb{R})$, since any measure belonging to a d -dimensional elementary face of $\mathcal{M}^+(\mathbb{T})$ is a sum of *exactly* d Dirac masses. Therefore we must have $\dim \mathcal{F}(m, \mathcal{M}^+(\mathbb{T})) < 2f_c + 1$ and, recalling (13), we deduce that the lower bound on $\text{codim}_{V \times \mathbb{R}}(\widehat{E} + \widehat{H})$ provided by Lemma 2.3 is far too pessimistic. In other words, the affine spaces determined by the Fourier coefficients only intersect very specific faces of the cone $\mathcal{M}^+(\mathbb{T})$.

An intuitive explanation consists in counting the “degrees of freedom” of the problem (we do not consider the statistical notion used in [22], but simply the “number of variables that should be fixed”). Informally, to fix the positions and amplitudes of k Dirac masses, that is $2k$ variables, we need at least $2k$ equations, i.e. $2k \leq 2f_c + 1$.

4.4. The basis-pursuit for measures

Surprisingly, things are different when considering the *Basis Pursuit for measures*,

$$\min_{m \in \mathcal{M}(\mathbb{T})} |m|(\mathbb{T}) \quad \text{s.t.} \quad \int_{\mathbb{T}} \varphi_k(t) dm(t) = y_k \quad (0 \leq k \leq 2f_c), \quad (28)$$

where $\{\varphi_k\}_{k=0}^{2f_c}$ is again the trigonometric system. In [9], Laurent Condat observed that when y is the Fourier coefficient vector of two opposite close spikes, a solution to (28) is a Dirac comb. More precisely, if $y = \Phi m_0 \stackrel{\text{def.}}{=} (\int_{\mathbb{T}} \varphi_k dm_0)_{k=0}^{2f_c}$ with $m_0 = \delta_{h/2} - \delta_{-h/2}$ and $0 < h < \frac{1}{2f_c}$, he noticed that the solution to (28) is given by

$$m = \sum_{j=-f_c}^{f_c-1} a_j \delta_{t_j}, \quad \text{where} \quad t_j \stackrel{\text{def.}}{=} \frac{1}{4f_c} + \frac{j}{2f_c}, \quad \text{and} \quad (29)$$

$$a_j = (-1)^j \frac{\cos(\pi h f_c)}{2f_c} \left(\cotan \left(\pi \left(\frac{1}{4f_c} + \frac{j}{2f_c} - \frac{h}{2} \right) \right) - \cotan \left(\pi \left(\frac{1}{4f_c} + \frac{j}{2f_c} + \frac{h}{2} \right) \right) \right). \quad (30)$$

While this observation in [9] seems to rely on numerical experiments, we provide in Section A a proof relying on a duality argument.

Proposition 4.1. *For $y = \Phi m_0$ with $m_0 = \delta_{h/2} - \delta_{-h/2}$ and $0 < h < 1/(2f_c)$, the unique solution to (28) is given by (29) and (30).*

As a consequence of Proposition 4.1, the number of Dirac masses predicted by Corollary 1.2 is *almost optimal* ($2f_c + 1$ Dirac masses are predicted whereas $2f_c$ actually appear). In fact, one cannot do “better”: it is proved in [9] that for every $y \in \mathbb{R}^{2f_c+1}$, there is a solution to (28) which is a sum of at most $2f_c$ Dirac masses.

Arguing informally in terms of “degrees of freedom”, we see that the above situation is quite peculiar: the relative positions of the Dirac masses are fixed, they can only move by a global translation. As a result, the $2f_c + 1$ variables determine the $2f_c$ amplitudes of the spikes and the last degree of freedom which is a global shift of the Dirac comb.

5. Conclusion

The representer theorem presented in this note describes, under very weak assumptions, the solutions of variational problems as a combination of a few atoms. It is optimal in the sense that for each described configuration, there is a variational problem for which the predicted number of atoms is attained. However, in some specific problems, the solutions may be sparser than predicted by the theorem. It is for instance the case of the truncated trigonometric moment problem involving the positivity constraint in the space of measures. On the other hand, with the total variation regularization for signed Radon measures, the prediction of the theorem is almost optimal.

Acknowledgements

The author warmly thanks Claire Boyer, Antonin Chambolle, Yohann De Castro, Frédéric de Gournay and Pierre Weiss for fruitful discussions about representer theorems. He also thanks an anonymous referee for stimulating comments about the connection between linear closedness and closedness in a topological sense.

This work was supported by the ANR CIPRESSI project, grant ANR-19-CE48-0017-01 of the French Agence Nationale de la Recherche.

A. Proof of Proposition 4.1

We endow $\mathcal{M}(\mathbb{T})$ with the weak* topology, and $\mathcal{C}(\mathbb{T})$ with the norm topology, so that $\mathcal{M}(\mathbb{T})$ and $\mathcal{C}(\mathbb{T})$ are *paired spaces* (in the sense of, e.g. [23]). A dual problem to (28) is

$$\sup_{p \in \mathbb{R}^{2f_c+1}} \int_{\mathbb{T}} \left(\sum_k p_k \varphi_k(t) \right) dm_0(t) \quad \text{s.t.} \quad \left\| \sum_k p_k \varphi_k \right\|_{\infty} \leq 1. \quad (31)$$

It can be shown that strong duality holds (see [12]) and that (31) has a solution. Moreover, any admissible measure m is a solution to (28) and p is a solution to (31) if and only if they satisfy the following extremality condition: the trigonometric polynomial $\eta \stackrel{\text{def.}}{=} \sum_k p_k \varphi_k$ satisfies $\eta(t) = 1$ for all $t \in \text{supp}(m_+)$ and $\eta(t) = -1$ for all $t \in \text{supp}(m_-)$, where m_+ and m_- denote the positive and negative parts in the Jordan decomposition of m . Therefore, it suffices to solve (31) to discover the possible support and sign of all the solutions to (28).

Lemma A.1. *For $0 < h < \frac{1}{2f_c}$, the unique solution to (31) is $p_* = (0, \dots, 0, 1)$, corresponding to the function $\eta_* : t \mapsto \sin(2\pi f_c t)$.*

Proof. As $m_0 = \delta_{h/2} - \delta_{-h/2}$, equation (31) can be reformulated as the maximization problem

$$\max_{\eta} (\eta(h/2) - \eta(-h/2)) \quad (32)$$

where the maximization is over all the trigonometric polynomials η with degree at most f_c and $\|\eta\|_{\infty} \leq 1$. Let η be a solution to (32). Note that its odd part, $\eta_{\text{odd}}(t) \stackrel{\text{def.}}{=} \frac{1}{2} (\eta(t) - \eta(-t))$, is also a solution to (32), so we first study η_{odd} .

Since $\|\eta_{\text{odd}}\|_{\infty} \leq 1$ and $\eta_*(t_j) = (-1)^j$ for $j \in \{-f_c, \dots, f_c - 1\}$, we note that $(-1)^j (\eta_* - \eta_{\text{odd}})(t_j) \geq 0$. Moreover,

$$\eta_{\text{odd}}(h/2) = \frac{1}{2} [\eta_{\text{odd}}(h/2) - \eta_{\text{odd}}(-h/2)] \geq \frac{1}{2} [\eta_*(h/2) - \eta_*(-h/2)] = \eta_*(h/2),$$

and by oddness $\eta_{\text{odd}}(-h/2) \leq \eta_*(-h/2)$.

As a result, if we define

$$\forall j \in \{-f_c - 2, \dots, f_c - 1\}, \quad t'_j \stackrel{\text{def.}}{=} \begin{cases} t_{j+2} & \text{for } -f_c \leq j \leq -3, \\ -h/2 & \text{for } j = -2, \\ h/2 & \text{for } j = -1, \\ t_j & \text{for } 0 \leq j \leq f_c - 1, \end{cases}$$

we have $2f_c + 2$ distinct points such that $(-1)^j(\eta_* - \eta_{\text{odd}})(t'_j) \geq 0$. Thus, $\eta_* - \eta_{\text{odd}}$ has at least $2f_c + 2$ roots (counting multiplicity): it is clear by the mean value theorem if each inequality is strict, and it can also be checked by approximation and counting the roots (with multiplicity) of the limit in the general case. As a consequence, $\eta_* - \eta_{\text{odd}}$ has at least $2f_c + 2$ roots and degree (at most) f_c : it must be identically zero.

We have proved that $\eta_{\text{odd}} = \eta_*$. Therefore it remains to deal with the even part: we write $\eta = \eta_* + \eta_{\text{even}}$, where η_{even} is the even part of η . Since $(-1)^j \eta(t_j) \leq 1$ for $-f_c \leq j \leq f_c - 1$, we get

$$1 + (-1)^j \eta_{\text{even}}(t_j) \leq 1,$$

so that $(-1)^j \eta_{\text{even}}(t_j) \leq 0$. Since $\eta_{\text{even}}(t_j) = \eta_{\text{even}}(t_{-j-1})$ we deduce that in fact $\eta_{\text{even}}(t_j) = 0$, so that η_{even} has $2f_c$ roots. But, since η and η_* reach their maximum or minimum at each t_j , we also have $0 = \eta'(t_j) = \eta'_*(t_j) + \eta'_{\text{even}}(t_j) = \eta'_{\text{even}}(t_j)$, so that t_j is a double root of η_{even} . Hence $\eta_{\text{even}} = 0$ and $\eta = \eta_*$. $\square$

In consequence, the support of every solution to (28) is contained in $\{t_{-f_c}, \dots, t_{f_c-1}\}$. Since that set is equispaced in $\mathbb{T}$, we notice as in [9] that, to recover the amplitudes a_j , we may invert the system

$$\sum_{j=-f_c}^{f_c-1} a_j e^{-2i\pi k t_j} = c_k \quad (-f_c \leq k \leq f_c - 1) \quad (33)$$

(where $c_0 = y_0$, $c_k = y_{2k-1} - i y_{2k}$ for $k \geq 1$ and $c_{-k} = c_k^*$ otherwise) using the Discrete Fourier Transform. We obtain

$$a_\ell = \frac{1}{2f_c} \sum_{k=-f_c}^{f_c-1} e^{2i\pi k (\frac{1}{4f_c} + \frac{\ell}{2f_c})} c_k.$$

Recalling that $c_k = e^{-2i\pi k \frac{h}{2}} - e^{2i\pi k \frac{h}{2}}$, we introduce the function

$$f(x) \stackrel{\text{def.}}{=} \frac{1}{2f_c} \sum_{k=-f_c}^{f_c-1} e^{2i\pi k x} = e^{-i\pi x} \frac{\sin(2\pi f_c x)}{2f_c \sin(\pi x)} = \frac{\sin(2\pi f_c x)}{2f_c} (\cotan(\pi x) - i), \quad (34)$$

so that $a_\ell = f(\frac{1}{4f_c} + \frac{\ell}{2f_c} - h/2) - f(\frac{1}{4f_c} + \frac{\ell}{2f_c} + h/2)$.

Since $\sin(\frac{\pi}{2} + \pi\ell + \pi h f_c) = \sin(\frac{\pi}{2} + \pi\ell - \pi h f_c) = (-1)^\ell \cos(\pi h f_c)$, we get

$$a_\ell = (-1)^\ell \frac{\cos(\pi h f_c)}{2f_c} \left(\cotan\left(\pi\left(\frac{1}{4f_c} + \frac{\ell}{2f_c} - \frac{h}{2}\right)\right) - \cotan\left(\pi\left(\frac{1}{4f_c} + \frac{\ell}{2f_c} + \frac{h}{2}\right)\right) \right).$$

Since the function $\cotan$ is (strictly) decreasing on $]0, \pi[$, we see that $a_\ell \neq 0$ with $\text{sign}(a_\ell) = \text{sign}(\eta_*(t_\ell))$, which is the desired optimality condition.

It remains to check that (33) also holds for $k = f_c$. Since $a_\ell \in \mathbb{R}$ for all ℓ , we obtain it by taking the conjugate of (33) for $k = -f_c$.

To summarize, there is only one measure m such that we have $(\int_{\mathbb{T}} \varphi_k dm(t))_k = y$ and $\eta_*(t) = 1$ for all $t \in \text{supp}(m_+)$ and $\eta_*(t) = -1$ for all $t \in \text{supp}(m_-)$. It is given by (29). The extremality conditions imply that it is the only solution to (28).

Remark A.2. For $h = \frac{1}{2f_c}$, a slight variation on the argument shows that the conclusions of Lemma A.1 still hold, and that the unique solution m satisfies $a_0 = 1$, $a_{-1} = -1$ and $a_\ell = 0$ otherwise. In other words, the measure $m_0 = \delta_{h/} - \delta_{-h/2}$ is perfectly recovered by (28). $\square$

References

- [1] A.I. Barvinok: *Problems of distance geometry and convex properties of quadratic maps*, Discrete Comp. Geometry 13/2 (1995) 189–202.
- [2] N. Boyd, G. Schiebinger, B. Recht: *The alternating descent conditional gradient method for sparse inverse problems*, SIAM J. Optimization 27/2 (2017) 616–639.
- [3] C. Boyer, A. Chambolle, Y. De Castro, V. Duval, F. de Gournay, P. Weiss: *On representer theorems and convex regularization*, SIAM J. Optimization 29/2 (2019) 1260–1281.
- [4] K. Bredies, M. Carioni: *Sparsity of solutions for variational inverse problems with finite-dimensional data*, Calc. Var. Partial Diff. Equations 59 (2020), art. no. 14.
- [5] K. Bredies, H. K. Pikkarainen: *Inverse problems in spaces of measures*, ESAIM: Control, Optimisation Calc. Variations 19/1 (2013) 190–218.
- [6] E. J. Candès, C. Fernandez-Granda: *Towards a mathematical theory of super-resolution*, Comm. Pure Appl. Math. 67/6 (2014) 906–956.
- [7] C. Carathéodory: *Über den Variabilitätsbereich der Koeffizienten von Potenzreihen, die gegebene Werte nicht annehmen*, Math. Annalen 64/1 (1907) 95–115.
- [8] V. Chandrasekaran, B. Recht, P. A. Parrilo, A. S. Willsky: *The convex geometry of linear inverse problems*, Found. Comp. Math. 12/6 (2012) 805–849.
- [9] L. Condat: *Atomic norm minimization for decomposition into complex exponentials and optimal transport in Fourier domain*, J. Approximation Theory 258 (2020).
- [10] Y. de Castro, F. Gamboa: *Exact reconstruction using Beurling minimal extrapolation*, J. Math. Analysis Appl. 395/1 (2012) 336–354.
- [11] L. E. Dubins: *On extreme points of convex sets*, J. Math. Analysis Appl. 5/2 (1962) 237–244.
- [12] V. Duval, G. Peyré: *Exact support recovery for sparse spikes deconvolution*, Found. Comp. Math. 15/5 (2015) 1315–1355.
- [13] S. D. Fisher, J. W. Jerome: *Spline solutions to L_1 extremal problems in one and several variables*, J. Approximation Theory 13/1 (1975) 73–83.
- [14] A. Flinth, P. Weiss: *Exact solutions of infinite dimensional total-variation regularized problems*, Inform. Inference: J. IMA 8/3 (2019) 407–443.

- [15] H. Gupta, J. Fageot, M. Unser: *Continuous-domain solutions of linear inverse problems with Tikhonov versus generalized TV regularization*, IEEE Trans. Signal Processing 66/17 (2018) 4670–4684.
- [16] V. Klee: *Convex sets in linear spaces*, Duke Math. J. 18/2 (1951) 443–466.
- [17] V. Klee: *Extremal structure of convex sets*, Archiv d. Mathematik 8/3 (1957) 234–240.
- [18] V. Klee: *On a theorem of Dubins*, J. Math. Analysis Appl. 7/3 (1963) 425–427.
- [19] M. G. Krein, 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*, Translations of Mathematical Monographs 50, American Mathematical Society, Providence (1977).
- [20] S. Lang: *Algebra*, rev. 3rd ed., Graduate Texts in Mathematics 211, Springer, New York (2002).
- [21] J. Matousek, B. Gärtner: *Understanding and Using Linear Programming*, Springer, Cham (2007).
- [22] C. Poon, G. Peyré: *Degrees of freedom for off-the-grid sparse estimation*, Bernoulli (2021).
- [23] R. T. Rockafellar: *Conjugate Duality and Optimization*, 2nd ed., Regional Conference Series in Applied Mathematics 16, SIAM, Philadelphia (1989).
- [24] R. T. Rockafellar: *Convex Analysis*, Princeton University Press, Princeton (1997).
- [25] O. Toeplitz: *Über die Fouriersche Entwicklung positiver Funktionen*, Rend. Circ. Mat. Palermo 32 (1911) 191–192.
- [26] M. Unser, J. Fageot, J. P. Ward: *Splines are universal solutions of linear inverse problems with generalized TV regularization*, SIAM Review 59/4 (2017) 769–793.
- [27] S. I. Zuhovickiĭ: *Remarks on problems in approximation theory (Ukrainian)*, Mat. Zbirnik KDU (1948) 169–183.