SIAM J. OPTIM.
Vol. 33, No. 2, pp. 842–867

THE GEOMETRY OF SPARSE ANALYSIS REGULARIZATION*

XAVIER DUPUIS† AND SAMUEL VAITER‡

Abstract. Analysis sparsity is a common prior in inverse problem or machine learning including special cases such as total variation regularization, edge Lasso, and fused Lasso. We study the geometry of the solution set (a polyhedron) of the analysis ℓ^1 regularization (with ℓ^2 data fidelity term) when it is not reduced to a singleton without any assumption of the analysis dictionary nor the degradation operator. In contrast with most theoretical work, we do not focus on giving uniqueness and/or stability results but rather describe a worst-case scenario where the solution set can be big in terms of dimension. Leveraging a fine analysis of the sublevel set of the regularizer itself, we draw a connection between support of a solution and the minimal face containing it and, in particular, prove that extreme points can be recovered thanks to an algebraic test. Moreover, we draw a connection between the sign pattern of a solution and the ambient dimension of the smallest face containing it. Finally, we show that any arbitrary subpolyhedra of the level set can be seen as a solution set of sparse analysis regularization with explicit parameters.

Key words. sparsity, geometry, uniqueness

MSC codes. 90C25, 49J52

DOI. 10.1137/19M1271877

1. Introduction.

We focus on a convex regularization problem promoting sparsity in an analysis dictionary in the context of a linear inverse problem/regression problem where the regularization reads

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|y-\Phi x\|_2^2+\lambda\|D^*x\|_1, \tag{1.1}$$

where $y\in\mathbb{R}^q$ is an observation/response vector, $\Phi\colon \mathbb{R}^n\to\mathbb{R}^q$ is the sensing/acquisition linear operator, $D\colon \mathbb{R}^p\to\mathbb{R}^n$ is a dictionary, and $\lambda>0$ the hyperparameter used as a trade-off between fidelity and regularization. Note that at this point, we do not make any assumption on the dictionary D or the acquisition operator Φ.

This convex regularization is known as analysis ℓ^1-regularization [11] in the inverse problems community or generalized Lasso [33] in statistics. Let us mention that it includes several popular regularizers as special cases as (anisotropic) total variation [23] when D is a discrete difference operator, wavelet coefficient analysis [28] using a wavelet transform as an analysis dictionary or fused Lasso [31] when using the concatenation of the identity matrix and a discrete difference operator, i.e., using a Lasso regularization with an additional constraint on the (discrete) gradient. In the noiseless context, when $y\in\operatorname{Im}\Phi$, the following constrained formulation is used instead of the Tikhonov formulation (1.1) as

$$\min_{x\in\mathbb{R}^n} \|D^*x\|_1 \quad \text{subject to} \quad \Phi x=y. \tag{1.2}$$

*Received by the editors July 1, 2019; accepted for publication (in revised form) December 5, 2022; published electronically June 20, 2023.
https://doi.org/10.1137/19M1271877
Funding: This work was partly supported by ANR GraVa ANR-18-CE40-0005 and Projet ANER RAGA G048CVCRB-2018ZZ.
†Université de Bourgogne, Dijon, France (xavier.dupuis@u-bourgogne.fr).
‡CNRS & Université Côte d'Azur, CNRS, LJAD, Nice 06000, France (samuel.vaiter@cnrs.fr).

We focus here on the noisy version of the regularization in order to keep our discussion concise. The purpose of this paper is to answer the following question:

> *When the solution set of* (1.1) *is not reduced to a singleton, what is its "geometry"?*

One possible motivation could be to study some *generalized solution path* of such a problem, not with respect to the hyperparameter λ (see, e.g., [10, 17, 33]) but with respect to some parameter of D. For example, consider

$$D_\rho = \begin{pmatrix} \rho & 0 \\ 0 & 1 \end{pmatrix}, \quad \Phi = \begin{pmatrix} 1 & 1 \end{pmatrix}, \quad y = 2, \quad \lambda = 1;$$

one can show that the solution set is

$$\{(1,0)\} \text{ if } \rho < 1, \quad [(1,0),(0,1)] \text{ if } \rho = 1, \quad \{(0,1)\} \text{ if } \rho > 1.$$

Even if most of the time the solution set is reduced to a singleton (see below), it is essential to describe what happens when it is not the case to understand the behavior of the generalized solution path. In this paper, we do not tackle fully this multivalued point of view and leave the sensitivity analysis for future work.

1.1. Previous works.

Uniqueness certificate of analysis regularization. Among several theoretical issues, sufficient condition for uniqueness of the solution set of (1.1) has been extensively studied; see, for instance, [36, 21, 1, 38, 29]. Several uniqueness conditions can be proposed, where the simplest is for instance requiring $n \leqslant q$ and Φ having full rank: the ℓ^2-loss term is strictly convex, and uniqueness follows from it. The task of studying the case when the solution set is not reduced to a singleton can be seen as rather formal since most of the time the solution set is reduced to a singleton [33, 36], but nevertheless, it exhibits interesting properties of sparse analysis regularization. See subsection 3.3 for a discussion of some of these conditions.

Solution set of generic convex program. Describing the geometry of the solution set in convex optimization has been a subject of intense study starting from the work of [19] and its generalization to nonsmooth convex programs [8]. Several extensions have been proposed such as [15] for pseudo-linear programs or in a different setting (minimization of concave function); [18] shows that one can describe one solution with minimal sparsity level.

Representer theorems. We shall also remark that coming from the statistics community, [16] (and popularized in [25]) initiates a line of work coined as representer theorems, culminating recently in [7] and [34]. The basic idea of these kinds of results is to show that under some assumptions, one can write every element of the solution set of a convex program as a sum of elementary atoms.

Description of polytopes. Convex polytopes and polyhedrons are central objects in geometry [39] and convex analysis. Part of our results provides a connection between faces and signs of vector living in the analysis domain. We can draw a connection with the study of oriented matroid [5] and zonotopes [6] (analysis ℓ^1-ball are zonotopes) as described in [39, Lecture 7], in particular in section 7.3.

1.2. Contributions.

In contrast to these lines of work, we take here a more direct and specific approach. We give below an overview of our contributions.

Geometry of the analysis ℓ^1-ball. The first part of our work (2) is dedicated to studying the geometry of the analysis ℓ^1-ball. We study across several results the direction and the relative interior of the intersection between the sublevel set of the regularizer and another set. We refine our analysis progressively starting from any convex component of the level set, then looking to subpolyhedra of the sublevel set ending by the faces itself of the level set. We show several specific results:

- The sign pattern defines a bijection between the set of exposed faces of the analysis ℓ^1-ball and the set of feasible signs in the dictionary D as proved in (2.24).
- The extreme points of the analysis ℓ^1-ball can be recovered with a purely algebraic result thanks to (2.25). We draw a link between our result and a remark in [7] which is a topological argument

Geometry of the solution set. Thanks to the study of the analysis ℓ^1-ball, we give in a second part (3) consequences on the solution set of (1.1). We show that

- Using (2.16) and (2.17), we describe the geometry of the solution set in (3.6).
- The solution set of (1.1) admits extreme points if and only if the condition denoted by (H_0) and assumed to hold all throughout [36] or [35], namely, $\operatorname{Ker}\Phi \cap \operatorname{Ker} D^* = \{0\}$, holds. In this case, the extreme points are precisely those which satisfy the condition denoted by (H_J) in [35] at a given solution to perform a sensitivity analysis; see (3.9).
- For any affine space which intersect nontrivially the unit-sphere, one can find Φ, y such that the solution set is exactly this intersection; see (3.11) and (3.12).

1.3. Notations. For a given integer n, the set of all integers between 1 and n is denoted by $[n] = \{1, \ldots, n\}$.

Vectors and support. Given $u \in \mathbb{R}^m$, the support $\operatorname{supp}(u)$ and the sign vector $\operatorname{sign}(u)$ are defined by

$$\operatorname{supp}(u) = \{i \in [m] \,:\, u_i \neq 0\} \quad \text{and} \quad \operatorname{sign}(u) = (\operatorname{sign}(u_i))_{i\in[m]},$$

and its cardinal is coined the ℓ^0-norm $\|u\|_0 = |\operatorname{supp}(u)|$. The cosupport $\operatorname{cosupp}(u)$ is the set $\operatorname{cosupp}(u) = [m]\setminus\operatorname{supp}(u)$. Given $u, v \in \mathbb{R}^m$, the inner product is written $\langle u,\, v\rangle = \sum_{i=1}^m u_i v_i$, and the associated norm is written $\|u\|_2 = \sqrt{\langle u,\, u\rangle}$. We will also use the ℓ^1-norm $\|u\|_1 = \sum_{i=1}^m |u_i|$ and ℓ^∞-norm $\|u\|_\infty = \max_{i\in[m]} |u_i|$.

Linear operators. Given a linear operator $D \in \mathbb{R}^{n\times m}$, $D^* \in \mathbb{R}^{m\times n}$ is the transpose operator, $D^+ \in \mathbb{R}^{n\times m}$ its Moore–Penrose pseudo-inverse, $\operatorname{Ker} D \subseteq \mathbb{R}^m$ its null space, and $\operatorname{Im} D \in \mathbb{R}^n$ its column space. Given $I \subseteq [m]$, $D_I \in \mathbb{R}^{n\times|I|}$ is the matrix formed by the column of D indexed by I. The identity operator is denoted Id_m or Id. Given a vector $u \in \mathbb{R}^m$, x_I is the vector of components indexed by I. Given a subspace $F \subseteq \mathbb{R}^m$, we denote by Π_F the orthogonal projection on F. Given a vector $u \in \mathbb{R}^n$, its diagonalized matrix $\operatorname{diag}(u) \in \mathbb{R}^{n\times n}$ is the diagonal matrix such that $\operatorname{diag}(u)_{ii} = u_i$ for every $i \in [n]$.

Convex analysis. Given a convex, lower semicontinuous, proper function $f : \mathbb{R}^m \to \mathbb{R}$, its subdifferential ∂f is given

$$\partial f(u) = \{\eta \in \mathbb{R}^m \,:\, f(u) \geqslant f(v) + \langle \eta,\, u - v\rangle\}\,.$$

Given a convex set C, the affine hull $\operatorname{aff}(C)$ is the smallest affine set containing C, the direction $\operatorname{dir}(C)$ of C is the direction of $\operatorname{aff}(C)$ and its relative interior $\operatorname{ri}(C)$ is the interior of C relative to its affine hull $\operatorname{aff}(C)$. The relative boundary $\operatorname{rbd}(C)$ of C is the boundary of C relative to $\operatorname{aff}(C)$. The dimension $\dim(C)$ of C is the dimension of $\operatorname{aff}(C)$. We say that $x \in C$ is an extreme point if there are no two different $x_1, x_2 \in C$ such that $x = \frac{x_1+x_2}{2}$. The set of all extreme points of C is denoted by $\mathrm{ext}(C)$. For instance, given two points $x_1 \neq x_2 \in \mathbb{R}^n$, the segment $C = [x_1, x_2]$ is such that its affine hull is $\mathrm{aff}(C) = \{x_1 + tx_2 \ : \ t \in \mathbb{R}\}$, its relative interior is the open segment $\mathrm{ri}(C) = (x_1, x_2)$, its relative boundary and set of extreme points $\mathrm{ext}(C) = \mathrm{rbd}(C) = \{x_1, x_2\}$, its dimension is 1, and its direction is $\mathrm{dir}(C) = \mathbb{R}(x_1 - x_2)$.

1.4. Examples of operators D^*. We illustrate our results in this paper in different analysis regularization settings. In particular, we focus our interest on different operators:

- The Lasso [30], corresponding to $D = D_{\text{Lasso}} = \mathrm{Id}$, used to recover sparse vectors.
- The total variation regularization [23], and more specifically the 1D total variation, i.e., when $D : \mathbb{R}^n \to \mathbb{R}^{n-1}$ is a forward difference operator on n points:

$$D^* = \begin{pmatrix} -1 & +1 & 0 & \cdots & 0 \\ 0 & -1 & +1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & +1 \end{pmatrix}.$$

  This is a popular prior in image processing to regularize "cartoon" or piecewise regular images.
- More generally, we consider the graph-total variation regularization [26] where $D = D_G$ is the vertex-edge incidence matrix of a graph G. Specific instance include the 1D and anisotropic 2D total variation [23], cluster Lasso [27].

2. The unit ball of the sparse analysis regularizer. This section contains the core of our results. After giving preliminary results on sign vectors in subsection 2.1, we show that the unit ball is a convex polyhedron by giving its half-space representation in subsection 2.2. Then, we study properties of convex subset of the unit-sphere in subsection 2.3 which leads us to Lemma 2.8 which turns to be the foundation of latter results. Subsection 2.4 contains a sequence of results which represent our main contribution: Theorem 2.16, which describes in detail the affine components of the unit ball, Proposition 2.17, which instantiates this result to the setting of an affine component included in the unit sphere, Proposition 2.18, which extends this result to any exposed faces, and finally Proposition 2.19, which gives a necessary and sufficient condition of extremality. Finally, in subsection 2.5, we reformulate our previous results in order to describe the exposed faces of the unit ball and to show that there exists a bijection between the set of exposed faces and feasible signs. We also draw a connection to the work of [7].

2.1. Preliminary results on sign vectors. We first define an order on the set of all possible signs $\{-1, 0, +1\}^p$ along with a notion of consistency of signs which can be related to the idea of "subsigns."

DEFINITION 2.1. *Let $s, s' \in \{-1, 0, +1\}^p$. We say that*

- *$s \preceq s'$ if for all $i \in [p]$, $s_i \neq 0 \ \Rightarrow s'_i = s_i$;*
- *s and s' are consistent if for all $i \in [p]$, $s_i \neq 0$ and $s'_i \neq 0 \ \Rightarrow s'_i = s_i$.*

The following remarks connect the notion of support/cosupport to this sign pattern.

Remark 2.2.

1. $s \preceq s' \Rightarrow \operatorname{supp}(s) \subset \operatorname{supp}(s') \Leftrightarrow \operatorname{cosupp}(s') \subset \operatorname{cosupp}(s)$;
2. If s and s' are consistent, then
$$s \preceq s' \Leftrightarrow \operatorname{supp}(s) \subset \operatorname{supp}(s') \Leftrightarrow \operatorname{cosupp}(s') \subset \operatorname{cosupp}(s);$$
3. If $s \preceq s''$ and $s' \preceq s''$, then s and s' are consistent;
4. The set $\{-1, 0, +1\}^p$ endowed with the order relation $\preceq$ is a poset.

The following lemma gives a characterization of the ℓ^1-norm which will be used intensively in latter results.

LEMMA 2.3. *Let $\theta \in \mathbb{R}^p$. Then for any $s \in \{-1, 0, +1\}^p$, $\langle s, \theta\rangle \leq \|\theta\|_1$, and the equality holds if and only if* $\operatorname{sign}(\theta) \preceq s$*.*

Proof. We prove the result componentwise. Let $\alpha \in \mathbb{R}$. Then for any $s \in \{-1, 0, +1\}$, $s\alpha \leq |\alpha|$. Suppose now that $\operatorname{sign}(\alpha) \preceq s$. If $\alpha = 0$, then $s\alpha = |\alpha|$, and if $\alpha \neq 0$, then $s = \operatorname{sign}(\alpha)$ and $s\alpha = |\alpha|$. Conversely, suppose that $\operatorname{sign}(\alpha) \npreceq s$. Then $\alpha \neq 0$ and $s \neq \operatorname{sign}(\alpha)$, i.e., $s = 0$ or $s = -\operatorname{sign}(\alpha)$. In both cases, $s\alpha < |\alpha|$. ∎

2.2. Half-space representation of the unit ball. We denote by B_1 (resp., ∂B_1) the unit ball (resp., the unit sphere) or sublevel set (resp., level set) for the value 1 of the sparse analysis regularizer $R\colon x \mapsto \|D^*x\|_1$:

$$B_1 = \{x \in \mathbb{R}^n : \|D^*x\|_1 \leq 1\},$$
$$\partial B_1 = \{x \in \mathbb{R}^n : \|D^*x\|_1 = 1\}.$$

Since R is one-homogeneous, the results of this section apply to all sublevel sets for positive values.

PROPOSITION 2.4. *The unit ball B_1 is a full-dimensional convex polyhedron, a half-space representation of which is given by*

$$B_1 = \bigcap_{s \in \{-1,0,1\}^p} \{x \in \mathbb{R}^n : \langle Ds, x\rangle \leq 1\}.$$

Proof. First note that B_1 has a nonempty interior (in particular $0 \in B_1$), namely, $\{x \in \mathbb{R}^n : \|D^*x\|_1 < 1\}$, which is equivalent for a convex set to be of full dimension.

Second denote $A = \bigcap_{s \in \{-1,0,1\}^p} \{x : \langle Ds, x\rangle \leqslant 1\}$. Let $x \in B_1$. By Lemma 2.3, $\langle Ds, x\rangle = \langle s, D^*x\rangle \leq \|D^*x\|_1 \leq 1$ for any $s \in \{-1, 0, 1\}^p$ so $x \in A$. Conversely, let $x \in A$ and $s = \operatorname{sign}(D^*x)$. By Lemma 2.3, $\|D^*x\|_1 = \langle s, D^*x\rangle = \langle Ds, x\rangle \leq 1$ so $x \in B_1$. Then $B_1 = A$ is a convex polyhedron. ∎

Note that this half-space representation is redundant, and if $D = \mathrm{Id}$, then it is the ℓ^1-ball. The general question of the minimal representation of H-polyhedron is known to be hard; we shall leave it to future work. However, we can use this proposition to derive a way to construct exposed face of B_1 as claimed in the following lemma.

LEMMA 2.5. *Let $\bar{s} \in \{-1, 0, 1\}^p$. Then*

$$B_1 \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\} = \partial B_1 \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\};$$

it is either empty or an exposed face of B_1.

Proof. Let $x \in B_1$. Then by Lemma 2.3, $\langle D\bar{s}, x\rangle = 1$ if and only if $1 = \langle \bar{s}, D^*x\rangle \leq \|D^*x\|_1 \leq 1$, if and only if $\|D^*x\|_1 = 1$ and $\operatorname{sign}(D^*x) \preceq \bar{s}$. If the intersection is nonempty, then $\{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}$ is a supporting hyperplane of B_1 by Proposition 2.4, and thus its intersection with B_1 is an exposed face of the polyhedron. ∎

For a given $\bar{s}$, the set $\{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\}$ looks hard to describe. In fact, there exists a linear representation as told in the following lemma.

LEMMA 2.6. *Let $\bar{s} \in \{-1, 0, 1\}^p$. Then*

$$\{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\} = \{x \in \mathbb{R}^n : D^*_{\bar{J}}x = 0 \text{ and } \operatorname{diag}(\bar{s}_{\bar{I}})D^*_{\bar{I}}x \geqslant 0\},$$

where $\bar{J} = \operatorname{cosupp}(\bar{s})$ and $\bar{I} = \operatorname{supp}(\bar{s})$.

Proof. It is a straightforward rewriting of $\operatorname{sign}(D^*x) \preceq \bar{s}$. Indeed,

$$\begin{aligned} \operatorname{sign}(D^*x) \preceq \bar{s} &\Leftrightarrow D^*_{\bar{J}}x = 0 \text{ and } \bar{s}_i(D^*x)_i \geqslant 0 \ \forall i \in \bar{I} \\ &\Leftrightarrow D^*_{\bar{J}}x = 0 \text{ and } \operatorname{diag}(\bar{s}_{\bar{I}})D^*_{\bar{I}}x \geqslant 0. \end{aligned}$$

∎

Note that we can exchange the role of $\bar{s}$ and D^*x, and we also obtain that

$$\{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\} = \{x \in \mathbb{R}^n : D^*_{\bar{J}}x = 0 \text{ and } \operatorname{diag}(D^*_{\bar{I}}x)\bar{s}_I \geqslant 0\}.$$

2.3. Convex components of the unit sphere. In this section we consider nonempty convex subsets $C \subset \partial B_1$. All the results will hold in particular for exposed faces of B_1.

We begin with a lemma on general convex sets.

LEMMA 2.7. *Let X be a nonempty convex set and $C \subset X$ be a nonempty convex subset. Suppose that there exists an exposed face G of X such that $\operatorname{ri}(C) \cap G \neq \emptyset$. Then $C \subset G$. Moreover, if G is exposed in $\operatorname{aff}(X)$, then $G \subset \operatorname{rbd}(X)$.*

Proof. Recall that an exposed face of X is defined as $G = X \cap \{x : \langle \alpha, x\rangle = \beta\}$ with $\{x : \langle \alpha, x\rangle = \beta\}$ a supporting hyperplane of X, i.e., such that $X \subset \{x : \langle \alpha, x\rangle \leq \beta\}$. Suppose that $C \not\subset G$, and let $x \in C \setminus G \subset X$ and $\bar{x} \in \operatorname{ri}(C) \cap G$ (nonempty). Then $\langle \alpha, x\rangle < \beta$ and $\langle \alpha, \bar{x}\rangle = \beta$. Let $d = \bar{x} - x \in \operatorname{dir}(C)$. Since $\bar{x} \in \operatorname{ri}(C)$, $\bar{x} + \varepsilon d \in C \subset X$ for $|\varepsilon|$ small. But $\langle \alpha, \bar{x} + \varepsilon d\rangle = \beta + \varepsilon(\beta - \langle \alpha, x\rangle) > \beta$ for $\varepsilon > 0$, which is a contradiction. Then $C \subset G$.

It is a classical result that $G \subset \operatorname{bd}(X)$; see, e.g., [14, Part III, section 2.4]. If G is exposed in $\operatorname{aff}(X)$, we get that $G \subset \operatorname{rbd}(X)$ by considering $\operatorname{aff}(X)$ as the ambient space. ∎

The following lemma is the first result of a long number of consequences which study the direction and relative interior of the intersection of the unit ball with another set.

LEMMA 2.8. *Let $\bar{x} \in \partial B_1$, $\bar{s} = \operatorname{sign}(D^*\bar{x})$, $\bar{J} = \operatorname{cosupp}(D^*\bar{x})$, and*

$$\bar{F} = B_1 \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}.$$

(i) $\bar{F} = \partial B_1 \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\}$ *and it is an exposed face;*
(ii) $C \subset \bar{F}$ *for any nonempty convex subset $C \subset B_1$ such that $\bar{x} \in \operatorname{ri}(C)$;*
(iii) $\operatorname{dir}(\bar{F}) = (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}}$;
(iv) $\operatorname{ri}(C) \subset \operatorname{ri}(\bar{F})$ *for any nonempty convex subset $C \subset B_1$ such that $\bar{x} \in \operatorname{ri}(C)$;*
(v) $\operatorname{ri}(\bar{F}) = \partial B_1 \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) = \bar{s}\}$.

Proof. (i) The expression for $\bar{F}$ is given by Lemma 2.5. It follows that $\bar{x} \in \bar{F}$, and thus $\bar{F}$ is an exposed face of B_1.

(ii) Let C be a nonempty convex subset of B_1 such that $\bar{x} \in \mathrm{ri}(C)$. Then $\mathrm{ri}(C) \cap \bar{F} \neq \emptyset$, and by Lemma 2.7, $C \subset \bar{F}$.

(iii) The inclusion $\mathrm{dir}(\bar{F}) \subset (D\bar{s})^\perp$ follows from the definition of $\bar{F}$. Moreover, for any $x \in \bar{F}$, $\mathrm{sign}(D^*x) \preceq \bar{s}$, which implies that $\bar{J} \subset \mathrm{cosupp}(D^*x)$, i.e., $D^*_{\bar{J}}x = 0$. Then $\bar{F} \subset \mathrm{Ker}\, D^*_{\bar{J}}$, and $\mathrm{dir}(\bar{F}) \subset \mathrm{Ker}\, D^*_{\bar{J}}$. Conversely, let $d \in (D\bar{s})^\perp \cap \mathrm{Ker}\, D^*_{\bar{J}}$. Since $d \in \mathrm{Ker}\, D^*_{\bar{J}}$ and $\mathrm{sign}(D^*\bar{x}) = \bar{s}$, $\mathrm{cosupp}(D^*\bar{x}) \subset \mathrm{cosupp}(D^*d)$ and $\mathrm{sign}(D^*(\bar{x}+\varepsilon d)) \preceq \bar{s}$ for $|\varepsilon|$ small. Then by Lemma 2.3, $\|\bar{x}+\varepsilon d\|_1 = \langle \bar{s}, D^*(\bar{x}+\varepsilon d)\rangle = \langle D\bar{s}, \bar{x}+\varepsilon d\rangle = \langle D\bar{s}, \bar{x}\rangle = 1$ since $d \in (D\bar{s})^\perp$. Then $\bar{x}+\varepsilon d \in \bar{F}$ for $|\varepsilon|$ small, and $d \in \mathrm{dir}(\bar{F})$.

(iv)–(v) First, we prove that $\partial B_1 \cap \{x \in \mathbb{R}^n : \mathrm{sign}(D^*x) = \bar{s}\} \subset \mathrm{ri}(\bar{F})$ (thus in particular $\bar{x} \in \mathrm{ri}(\bar{F})$). Let $x \in \partial B_1$ be such that $\mathrm{sign}(D^*x) = \bar{s}$. By (iii) and its proof, for any $d \in \mathrm{dir}(\bar{F})$, $x + \varepsilon d \in \bar{F}$ for $|\varepsilon|$ small. Thus $x \in \mathrm{ri}(\bar{F})$. Second, let C be a nonempty convex subset of B_1 such that $\bar{x} \in \mathrm{ri}(C)$. By (ii), $C \subset \bar{F} \subset \partial B_1$. Let $\hat{x} \in \mathrm{ri}(C)$ and $\hat{s} = \mathrm{sign}(D^*\hat{x})$. Note that $\hat{s} \preceq \bar{s}$ since $\hat{x} \in \bar{F}$. Let $\hat{F} = B_1 \cap \{x \in \mathbb{R}^n : \langle D\hat{s}, x\rangle = 1\}$. Applying (ii) to $\hat{x}$ and $C = \bar{F}$, we get that $\bar{F} \subset \hat{F}$, which implies that $\bar{s} \preceq \hat{s}$. Thus $\mathrm{sign}(D^*\hat{x}) = \bar{s}$ and $\mathrm{ri}(C) \subset \partial B_1 \cap \{x \in \mathbb{R}^n : \mathrm{sign}(D^*x) = \bar{s}\}$. This proves (iv) as well as the missing inclusion of (v) by setting $C = \bar{F}$ (recall that $\bar{x} \in \mathrm{ri}(\bar{F})$). $\square$

The following proposition is a direct consequence of Lemma 2.8 which allows to characterize faces of B_1 by an arbitrary convex subset of it.

PROPOSITION 2.9. *Let C be a nonempty convex subset of ∂B_1. Let $\bar{x} \in \mathrm{ri}(C)$, $\bar{s} = \mathrm{sign}(D^*\bar{x})$, and*

$$\bar{F} = B_1 \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}.$$

Then $C \subset \bar{F}$ and $\mathrm{ri}(C) \subset \mathrm{ri}(\bar{F})$. Moreover, $\bar{F}$ is the smallest face of B_1 such that $\mathrm{ri}(C) \cap \bar{F} \neq \emptyset$ and the unique face of B_1 such that $\mathrm{ri}(C) \cap \mathrm{ri}(\bar{F}) \neq \emptyset$.

Proof. By Lemma 2.8(i) and (iii), $C \subset \bar{F}$ and $\mathrm{ri}(C) \subset \mathrm{ri}(\bar{F})$. Let F be a face such that $\mathrm{ri}(C) \cap F \neq \emptyset$. If $F = B_1$; then $\bar{F} \subset F$; otherwise F is an exposed face since B_1 is a polyhedron, and by Lemma 2.7, $C \subset F$. It follows that $\emptyset \neq \mathrm{ri}(C) \cap \mathrm{ri}(\bar{F}) \subset F \cap \mathrm{ri}(\bar{F})$, and by Lemma 2.7 again, $\bar{F} \subset F$. Suppose now that $\mathrm{ri}(C) \cap \mathrm{ri}(F) \neq \emptyset$. Then permuting F and $\bar{F}$, we get that $F \subset \bar{F}$, thus $F = \bar{F}$. $\square$

Remark 2.10. The uniqueness actually holds with the same proof for a general nonempty convex set X: given a nonempty convex subset $C \subset \mathrm{rbd}(X)$, there exists at most one exposed face G of X such that $\mathrm{ri}(C) \cap \mathrm{ri}(G) \neq \emptyset$. The existence reduces to the existence, for any $x \in \mathrm{rbd}(X)$, of an exposed face G of X such that $x \in \mathrm{ri}(G)$.

For a singleton $C = \{\bar{x}\}$, the previous proposition becomes the following.

COROLLARY 2.11. *Let $\bar{x} \in \partial B_1$ and $\bar{s} = \mathrm{sign}(D^*\bar{x})$. Then $\bar{F} = B_1 \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}$ is the smallest face of B_1 such that $\bar{x} \in \bar{F}$ and the unique face of B_1 such that $\bar{x} \in \mathrm{ri}(\bar{F})$.*

We can also derive from Proposition 2.9 and Lemma 2.8 the following properties about the mapping $x \mapsto \mathrm{sign}(D^*x)$.

COROLLARY 2.12. *Let C be a nonempty convex subset of the unit sphere ∂B_1. Then $\max_{x \in C} \mathrm{sign}(D^*x)$ is well defined, and this maximum is attained everywhere in $\mathrm{ri}(C)$. In particular $\mathrm{sign}(D^*\cdot)$ is constant on $\mathrm{ri}(C)$.*

For any $x, x' \in C$, $[x, x']$ is a nonempty convex subset of ∂B_1 and thus $\mathrm{sign}(D^*\cdot)$ is constant on $]x, x'[$. Moreover, since $s = \mathrm{sign}(D^*x)$ and $s' = \mathrm{sign}(D^*x')$ are both $\preceq \max_{x \in C} \mathrm{sign}(D^*x)$, they are consistent (see Remark 2.2). It follows that the constant value s'' of $\mathrm{sign}(D^*\cdot)$ on $]x, x'[$ can be given explicitly:

$$s''_i = \begin{cases} s_i & \text{if } s_i \neq 0, \\ s'_i & \text{if } s'_i \neq 0, \\ 0, & \text{otherwise.} \end{cases}$$

It is also the maximum of $\operatorname{sign}(D^*\cdot)$ over $[x, x']$.

Finally we get a general sufficient condition of extremality.

COROLLARY 2.13. *Let C be a nonempty convex subset of ∂B_1. Let $\bar{x} \in C$ and $\bar{s} = \operatorname{sign}(D^*\bar{x})$. If $\bar{x}$ is the unique $x \in C$ such that $\operatorname{sign}(D^*x) \preceq \bar{s}$, then $\bar{x} \in \operatorname{ext}(C)$.*

Proof. Let $x_1, x_2 \in C$ such $\bar{x} = \frac{x_1+x_2}{2}$. Then $[x_1, x_2]$ is a nonempty convex subset of ∂B_1. By Corollary 2.12, $\operatorname{sign}(D^*x) = \bar{s}$ for any $x \in]x_1, x_2[$. By uniqueness of $\bar{x}$, $]x_1, x_2[= \{\bar{x}\}$, i.e., $x_1 = x_2 = \bar{x}$, and thus $\bar{x}$ is an extreme point. ∎

Observe that the uniqueness condition in the Corollary 2.13 can be written as $C \cap \bar{F} = \{\bar{x}\}$ with $\bar{F} = B_1 \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}$ the smallest face of B_1 containing $\bar{x}$.

2.4. Subpolyhedra of the unit ball. In this section we consider nonempty convex polyhedra of the form $\mathcal{A} \cap B_1$ with $\mathcal{A}$ an affine subspace. The results on convex components of the unit sphere apply to such sets if $\mathcal{A} \cap B_1 \subset \partial B_1$ and, in any case, as we will see, to exposed faces of such polyhedra. Again, the results of this section will hold in particular for exposed faces of B_1.

We begin with a useful lemma.

LEMMA 2.14. *Let $\mathcal{A}$ be an affine subspace and C be a nonempty convex set such that $\mathcal{A} \cap \operatorname{ri}(C) \neq \emptyset$. Then*

$$\operatorname{ri}(\mathcal{A} \cap C) = \mathcal{A} \cap \operatorname{ri}(C) \textit{ and } \operatorname{dir}(\mathcal{A} \cap C) = \operatorname{dir}(\mathcal{A}) \cap \operatorname{dir}(C).$$

Proof. Since $\operatorname{ri}(\mathcal{A}) = \mathcal{A}$, we have $\operatorname{ri}(\mathcal{A}) \cap \operatorname{ri}(C) \neq \emptyset$. Then by [14, Part III, Proposition 2.1.10], $\operatorname{ri}(\mathcal{A} \cap C) = \operatorname{ri}(\mathcal{A}) \cap \operatorname{ri}(C) = \mathcal{A} \cap \operatorname{ri}(C)$. Let us now prove that $\operatorname{dir}(\mathcal{A} \cap C) = \operatorname{dir}(\mathcal{A}) \cap \operatorname{dir}(C)$. Let $d \in \operatorname{dir}(\mathcal{A} \cap C)$ and $\bar{x} \in \operatorname{ri}(\mathcal{A} \cap C)$ (nonempty). Then $\bar{x} + \varepsilon d \in \mathcal{A} \cap C$ for $|\varepsilon|$ small, and $d \in \operatorname{dir}(\mathcal{A}) \cap \operatorname{dir}(C)$. Similarly, let $d \in \operatorname{dir}(\mathcal{A}) \cap \operatorname{dir}(F)$ and $\bar{x} \in \operatorname{ri}(\mathcal{A}) \cap \operatorname{ri}(C)$ (nonempty). Then $\bar{x} + \varepsilon d \in \mathcal{A} \cap C$ for $|\varepsilon|$ small, and $d \in \operatorname{dir}(\mathcal{A} \cap C)$. ∎

The following lemma will be used in Theorem 2.16.

LEMMA 2.15. *Let $\mathcal{A}$ be an affine subspace such that $\emptyset \neq \mathcal{A} \cap B_1 \subset \partial B_1$. Let $\bar{x} \in \mathcal{A} \cap \partial B_1$, $\bar{s} = \operatorname{sign}(D^*\bar{x})$, and $\bar{J} = \operatorname{cosupp}(D^*\bar{x})$. Then*

(i) $\operatorname{dir}(\mathcal{A}) \subset \bigcup_{s \succeq \bar{s}} \{d \in \mathbb{R}^n : \langle Ds, d\rangle \geq 0\}$;

(ii) $\operatorname{dir}(\mathcal{A})^\perp \cap (\sum_{s \succeq \bar{s}} \mathbb{R}_+ Ds) \neq \{0\}$;

(iii) $\operatorname{dir}(\mathcal{A}) \cap \operatorname{Ker} D^*_{\bar{J}} \subset (D\bar{s})^\perp$, *i.e.,* $\operatorname{dir}(\mathcal{A}) \cap (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}} = \operatorname{dir}(\mathcal{A}) \cap \operatorname{Ker} D^*_{\bar{J}}$.

(iv) $\mathcal{A} \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\} \subset \partial B_1$, *i.e.,* $\mathcal{A} \cap \partial B_1 \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\} = \mathcal{A} \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\}$;

(v) $\mathcal{A} \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) = \bar{s}\} \subset \partial B_1$, *i.e.,* $\mathcal{A} \cap \partial B_1 \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) = \bar{s}\} = \mathcal{A} \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) = \bar{s}\}$.

Proof. (i) Suppose that the inclusion does not hold, and let $d \in \operatorname{dir}(\mathcal{A})$ such that $\langle Ds, d\rangle < 0$ for all $s \succeq \bar{s}$. Then $\langle Ds, \bar{x} + \varepsilon d\rangle < 1$ for all s and $\varepsilon > 0$ small. Indeed, by Lemma 2.3, if $s \succeq \bar{s}$, then $\langle Ds, \bar{x} + \varepsilon d\rangle < 1$ for $\varepsilon > 0$; otherwise, $\langle Ds, \bar{x} + \varepsilon d\rangle < 1$ for ε small. Still by Lemma 2.3, $\|\bar{x} + \varepsilon d\|_1 < 1$ for $\varepsilon > 0$ small; i.e., $\bar{x} + \varepsilon d \in \mathcal{A} \cap \mathring{B}_1$, which is in contradiction with $\mathcal{A} \cap B_1 \subset \partial B_1$.

(ii) Suppose that the intersection is reduced to $\{0\}$ and consider the dual cone of both sides of the expression (we denote by $C^* = \{y \in \mathbb{R}^n : \forall x \in C, \langle y, x\rangle \geq 0\}$ of a subset $C \subset \mathbb{R}^n$). Since $\operatorname{dir}(\mathcal{A})^\perp$ and $(\sum_{s\succeq\bar{s}} \mathbb{R}_+ Ds) \neq \{0\}$ are two polyhedral cones, we get that

$$\operatorname{dir}(\mathcal{A})^{\perp *} + \Big(\sum_{s\succeq\bar{s}} \mathbb{R}_+ Ds\Big)^* = \mathbb{R}^n,$$

where $\operatorname{dir}(\mathcal{A})^{\perp *} = \operatorname{dir}(\mathcal{A})$ and $(\sum_{s\succeq\bar{s}} \mathbb{R}_+ Ds)^* = \bigcap_{s\succeq\bar{s}} Ds^*$. It follows from (i) that $\mathbb{R}^n \subset \bigcup_{s\succeq\bar{s}} Ds^*$, which is not true (note, e.g., that $d = x - \bar{x}$ with $x \in \mathring{B}_1$ is such that $\langle Ds, d\rangle < 0$) for all $s \succeq \bar{s}$).

(iii) First note that for $d \in \operatorname{Ker} D^*_{\bar{J}}$, $\langle Ds, d\rangle = \langle D\bar{s}, d\rangle$ for all $s \succeq \bar{s}$. Indeed, $(D^*d)_i = 0$ if $\bar{s}_i = 0$ and $s_i = \bar{s}_i$ if $\bar{s}_i \neq 0$; thus $s_i(D^*d)_i = \bar{s}_i(D^*d)_i$ for all i and $\langle s, D^*d\rangle = \langle \bar{s}, D^*d\rangle$. Let now $d \in \operatorname{dir}(\mathcal{A}) \cap \operatorname{Ker} D^*_{\bar{J}}$. Since by (i) there exists $s \succeq \bar{s}$ such that $\langle Ds, d\rangle \geq 0$, it follows that $\langle D\bar{s}, d\rangle \geq 0$. And since $-d \in \operatorname{dir}(\mathcal{A}) \cap \operatorname{Ker} D^*_{\bar{J}}$ too, we get that $\langle D\bar{s}, d\rangle = 0$, which proves the inclusion and the equivalent equality.

(iv) Let $x \in \mathcal{A}$ such that $\operatorname{sign}(D^*x) \preceq \bar{s}$ (in particular, $\bar{J} \subset \operatorname{cosupp}(D^*x)$), and let $d = x - \bar{x}$. Then $d \in \operatorname{dir}(\mathcal{A}) \cap \operatorname{Ker} D^*_{\bar{J}}$ (recall that $\bar{x} \in \mathcal{A}$ and $\bar{J} = \operatorname{cosupp}(D^*\bar{x})$), and by (ii), $d \in (D\bar{s})^\perp$. By Lemma 2.3, $\|D^*x\|_1 = \langle \bar{s}, D^*x\rangle = \langle D\bar{s}, \bar{x} + d\rangle = \langle D\bar{s}, \bar{x}\rangle = \|D^*\bar{x}\|_1 = 1$, which proves the inclusion and the equivalent equality.

(v) The proof is the same as for (iv). ∎

The following theorem is similar to Lemma 2.8 when we replace convex subset by subpolyhedra (here of the unit sphere).

Theorem 2.16. *Let $\mathcal{A}$ be an affine subspace intersecting ∂B_1. Let $\bar{x} \in \mathcal{A} \cap \partial B_1$, $\bar{s} = \operatorname{sign}(D^*\bar{x})$, $\bar{J} = \operatorname{cosupp}(D^*\bar{x})$, and $\bar{F} = B_1 \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}$. Then*

$$\bar{G} = \mathcal{A} \cap \bar{F}$$

(i) *is the smallest face of $\mathcal{A} \cap B_1$ such that $\bar{x} \in \bar{G}$ and the unique face of $\mathcal{A} \cap B_1$ such that $\bar{x} \in \operatorname{ri}(\bar{G})$ ($\bar{G}$ is possibly equal to $\mathcal{A} \cap B_1$ itself);*

(ii) *satisfies the following:*

$$\begin{aligned}
\bar{G} &= \mathcal{A} \cap \partial B_1 \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\},\\
\operatorname{ri}(\bar{G}) &= \mathcal{A} \cap \partial B_1 \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) = \bar{s}\},\\
\operatorname{dir}(\bar{G}) &= \operatorname{dir}(\mathcal{A}) \cap (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}};
\end{aligned}$$

(iii) *satisfies the following, in the case where $\mathcal{A} \cap B_1 \subset \partial B_1$:*

$$\begin{aligned}
\bar{G} &= \mathcal{A} \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\},\\
\operatorname{ri}(\bar{G}) &= \mathcal{A} \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) = \bar{s}\},\\
\operatorname{dir}(\bar{G}) &= \operatorname{dir}(\mathcal{A}) \cap \operatorname{Ker} D^*_{\bar{J}}.
\end{aligned}$$

Proof. (i) First note that $\bar{G} = (\mathcal{A} \cap B_1) \cap \{x : \langle D\bar{s}, x\rangle = 1\}$ is an exposed face of $\mathcal{A} \cap B_1$ (in particular it is a convex subset of $\mathcal{A} \cap B_1$) and is such that $\bar{x} \in \operatorname{ri}(\bar{G}) = \mathcal{A} \cap \operatorname{ri}(\bar{F})$ by Lemma 2.14. Let G be a face of $\mathcal{A} \cap B_1$ such that $\bar{x} \in G$. If $G = \mathcal{A} \cap B_1$, then $\bar{G} \subset G$; otherwise G is an exposed face since $\mathcal{A} \cap B_1$ is a convex polyhedron, and by Lemma 2.7, $\bar{G} \subset G$. Suppose now that $\bar{x} \in \operatorname{ri}(G)$. Then permuting G and $\bar{G}$, we get that $G \subset \bar{G}$, thus $G = \bar{G}$.

(ii) By definition, $\bar{G} = \mathcal{A} \cap \bar{F}$ and $\bar{x} \in \mathcal{A} \cap \operatorname{ri}(\bar{F})$. By Lemma 2.14, $\operatorname{ri}(\bar{G}) = \mathcal{A} \cap \operatorname{ri}(\bar{F})$ and $\operatorname{dir}(\bar{G}) = \operatorname{dir}(\mathcal{A}) \cap \operatorname{dir}(\bar{F})$. The expression of these sets follows from Lemma 2.8.

(iii) We proved the strengthened expression of the previous sets in Lemma 2.15. ∎

We get the next result on $\mathcal{A} \cap B_1$ itself in the case where it is a subset of ∂B_1.

PROPOSITION 2.17. *Let $\mathcal{A}$ be an affine subspace such that $\emptyset \neq \mathcal{A} \cap B_1 \subset \partial B_1$. Then*

$$\mathcal{A} \cap B_1 = \mathcal{A} \cap \bar{F}$$

with $\bar{F} = B_1 \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}$ and $\bar{s} = \max_{x \in \mathcal{A}\cap B_1} \operatorname{sign}(D^*x)$ (or equivalently $\bar{s} = \operatorname{sign}(D^*\bar{x})$ for some $\bar{x} \in \operatorname{ri}(\mathcal{A} \cap B_1)$). In particular, the results of Theorem 2.16 (iii) *hold for $\mathcal{A} \cap B_1$.*

Proof. Since $\mathcal{A} \cap B_1$ is a nonempty convex subset of ∂B_1, $\bar{s}$ is well defined by Corollary 2.12. Let $\bar{x} \in \operatorname{ri}(\mathcal{A} \cap B_1) \subset \mathcal{A} \cap \partial B_1$. Then by Theorem 2.16(i), $\mathcal{A} \cap \bar{F}$ is the unique face of $\mathcal{A} \cap B_1$ containing $\bar{x}$ in its relative interior; it is thus equal to the face $\mathcal{A} \cap B_1$. ∎

In the general case, we can describe all the exposed faces of $\mathcal{A} \cap B_1$.

PROPOSITION 2.18. *Let $\mathcal{A}$ be an affine subspace such that $\emptyset \neq \mathcal{A} \cap B_1 \neq \operatorname{aff}(\mathcal{A} \cap B_1)$. Let G be a face $\mathcal{A} \cap B_1$ exposed in $\operatorname{aff}(\mathcal{A} \cap B_1)$. Then*

$$G = \mathcal{A} \cap \bar{F}$$

with $\bar{F} = B_1 \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}$ and $\bar{s} = \max_{x \in G} \operatorname{sign}(D^*x)$ (or equivalently $\bar{s} = \operatorname{sign}(D^*\bar{x})$ for some $\bar{x} \in \operatorname{ri}(G)$). In particular, the results of Theorem 2.16(ii) *(or* (iii) *if $\mathcal{A} \cap B_1 \subset \partial B_1$) hold for G.*

Proof. By Lemma 2.7, $G \subset \operatorname{rbd}(\mathcal{A} \cap B_1)$. Let us show that $\operatorname{rbd}(\mathcal{A} \cap B_1) \subset \mathcal{A} \cap \partial B_1$. We distinguish two cases: if $\mathcal{A} \cap B_1 \subset \partial B_1$, there is nothing to prove; otherwise, $\mathcal{A} \cap \mathring{B}_1 \neq \emptyset$. By Lemma 2.14 applied the full-dimensional convex B_1, $\operatorname{ri}(\mathcal{A} \cap B_1) = \operatorname{ri}(\mathcal{A}) \cap \operatorname{ri}(B_1) = \mathcal{A} \cap \mathring{B}_1$. Then $\operatorname{rbd}(\mathcal{A} \cap B_1) = \mathcal{A} \cap B_1 \setminus \mathcal{A} \cap \mathring{B}_1 = \mathcal{A} \cap \partial B_1$. In particular, G is a nonempty convex subset of ∂B_1, thus $\bar{s}$ is well defined by Corollary 2.12.

Let $\bar{x} \in \operatorname{ri}(G) \subset \mathcal{A} \cap \partial B_1$. Then by Theorem 2.16(i), $\mathcal{A} \cap \bar{F}$ is the unique face of $\mathcal{A} \cap B_1$ containing $\bar{x}$ in its relative interior; it is thus equal to the face G. ∎

In this setting, we get a necessary and sufficient condition of extremality. Note that the notion of extremality can be related to the topology of the set; here our condition only uses an algebraic characterization.

PROPOSITION 2.19. *Let $\mathcal{A}$ be an affine subspace intersecting ∂B_1. Let $\bar{x} \in \mathcal{A} \cap \partial B_1$, $\bar{s} = \operatorname{sign}(D^*\bar{x})$, and $\bar{J} = \operatorname{cosupp}(D^*\bar{x})$. Then*

$$\begin{aligned} \bar{x} \in \operatorname{ext}(\mathcal{A} \cap B_1) &\Leftrightarrow \operatorname{dir}(\mathcal{A}) \cap (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}} = \{0\} \\ &\Leftrightarrow \operatorname{dir}(\mathcal{A}) \cap \operatorname{Ker} D^*_{\bar{J}} = \{0\} \text{ in the case where } \mathcal{A} \cap B_1 \subset \partial B_1. \end{aligned}$$

Moreover, $\mathcal{A} \cap B_1$ admits extreme points (that necessarily belong to $\mathcal{A} \cap \partial B_1$) if and only if it is compact (i.e., is a convex polytope) if and only if $\operatorname{dir}(\mathcal{A}) \cap \operatorname{Ker} D^* = \{0\}$.

Proof. Since $\mathcal{A} \cap B_1$ is a convex polyhedron, $\bar{x} \in \operatorname{ext}(\mathcal{A} \cap B_1) \Leftrightarrow \{\bar{x}\}$ is a face $\mathcal{A} \cap B_1 \Leftrightarrow \{\bar{x}\} = \bar{G}$ of Theorem 2.16 $\Leftrightarrow \operatorname{dir}(\bar{G}) = \{0\}$.

Recall that extreme points belong to $\operatorname{rbd}(\mathcal{A} \cap B_1)$ and that, as in Proposition 2.18, $\operatorname{rbd}(\mathcal{A} \cap B_1) \subset \mathcal{A} \cap \partial B_1$. Note also that we always have $\operatorname{dir}(\mathcal{A}) \cap \operatorname{Ker} D^* \subset \operatorname{dir}(\mathcal{A}) \cap (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}}$. Thus if $\mathcal{A} \cap B_1$ admits extreme points, then $\operatorname{dir}(\mathcal{A}) \cap \operatorname{Ker} D^* = \{0\}$ by the beginning of the corollary, which implies that $\mathcal{A} \cap B_1$ is bounded and thus compact, which in turn implies the existence of extreme points [14, Part III, Proposition 2.3.3]. ∎

Remark 2.20. It is possible to show directly (via Corollary 2.13) that the condition above is sufficient. Indeed, assume that $\operatorname{dir}(\mathcal{A}) \cap (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}} = \{0\}$. Let $x \in$

$\mathcal{A} \cap \partial B_1$ such that $\operatorname{sign}(D^* x) \preceq \bar{s}$; let us show that $x = \bar{x}$: $x - \bar{x} \in \operatorname{dir}(\mathcal{A})$; by Lemma 2.3, $\langle \bar{s}, D^* x\rangle = \|D^* x\|_1 = \|D^* \bar{x}\|_1 = \langle \bar{s}, D^* \bar{x}\rangle$, thus $x - \bar{x} \in (D\bar{s})^\perp$; by Remark 2.2, $\bar{J} \subset \operatorname{cosupp}(D^* x)$, thus $x - \bar{x} \in \operatorname{Ker} D^*_{\bar{J}}$. Then $\bar{s}$ is minimal, and by Corollary 2.13, $\bar{x} \in \operatorname{ext}(\mathcal{A} \cap B_1)$.

2.5. Consequence results on the unit ball. The previous results will be at the core of our study of the solution set of (1.1) in section 3. Nevertheless, we can also dive deeper into this analysis in order to fully characterize the faces of the unit-ball as a byproduct.

A first consequence or reformulation of the previous results is that all the exposed faces of B_1 are of the form $B_1 \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}$.

PROPOSITION 2.21. *Let F be an exposed face of B_1. Then*

$$F = B_1 \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}$$

with $\bar{s} = \max_{x \in F} \operatorname{sign}(D^* x)$ (or equivalently $\bar{s} = \operatorname{sign}(D^* \bar{x})$ for some $\bar{x} \in \operatorname{ri}(F)$). Moreover,

$$\begin{aligned}
F &= \partial B_1 \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^* x) \preceq \bar{s}\} \\
&= \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\} \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^* x) \preceq \bar{s}\}, \\
\operatorname{ri}(F) &= \partial B_1 \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^* x) = \bar{s}\} \\
&= \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\} \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^* x) = \bar{s}\}, \\
\operatorname{dir}(F) &= (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}} \text{ (where } \bar{J} = \operatorname{cosupp}(\bar{s})).
\end{aligned}$$

Proof. Since F is a nonempty convex subset of\ ∂B_1, $\bar{s}$ is well defined by Corollary 2.12. The first statement is a consequence of Corollary 2.11. The next statements follow from Lemma 2.8 and Proposition 2.17 with $\mathcal{A} = \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}$ a supporting hyperplane of B_1 defining the face F (recall or note by Lemma 2.7 that $\mathcal{A} \cap B_1 \subset \partial B_1$ for any $\mathcal{A}$ supporting hyperplane of B_1). ∎

Remark 2.22. Any exposed face F of B_1 satisfies $\operatorname{Ker} D^* \subset \operatorname{dir}(F) \subset (D\bar{s})^\perp$ with equality if and only if $(D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}} = \operatorname{Ker} D^*$ (for the left inclusion) and $(D\bar{s})^\perp \subset \operatorname{Ker} D^*_{\bar{J}}$ (for the right inclusion).

Together with Lemma 2.6, the previous proposition gives the following half-space representations of the exposed faces of B_1.

COROLLARY 2.23. *Let F be an exposed face of B_1. Then*

$$F = \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1, D^*_{\bar{J}} x = 0, \operatorname{diag}(\bar{s}_{\bar{I}}) D^*_{\bar{I}} x \geq 0\}$$

with $\bar{s} = \max_{x \in F} \operatorname{sign}(D^* x)$, $\bar{J} = \operatorname{cosupp}(\bar{s})$, and $\bar{I} = \operatorname{supp}(\bar{s})$.

By Lemma 2.8 and Proposition 2.21, the mapping $F \mapsto \max_{x \in F} \operatorname{sign}(D^* x)$ is a bijection between the set of exposed faces of B_1 and the set of feasible signs $\{\operatorname{sign}(D^* x) : x \in \partial B_1\}$. The next observation is that this bijection preserves the partial orders (i.e., it is an order isomorphism).

PROPOSITION 2.24. *Let F_1 and F_2 be two exposed faces of B_1 and*

$$s_i = \max_{x \in F_i} \operatorname{sign}(D^* x), \ i = 1, 2.$$

Then

$$F_1 \subset F_2 \Leftrightarrow s_1 \preceq s_2.$$

In this case, denoting by $J_1 = \operatorname{cosupp}(s_1)$,

$$\begin{aligned} F_1 &= F_2 \cap \{x \in \mathbb{R}^n : J_1 \subset \operatorname{cosupp}(D^* x)\}, \\ \operatorname{ri}(F_1) &= F_2 \cap \{x \in \mathbb{R}^n : J_1 = \operatorname{cosupp}(D^* x)\}, \\ \operatorname{dir}(F_1) &= \operatorname{dir}(F_2) \cap \operatorname{Ker} D^*_{J_1}. \end{aligned}$$

Proof. Suppose that $F_1 \subset F_2$, and let $x \in \operatorname{ri}(F_1) \subset F_2$; then $s_1 = \operatorname{sign}(D^* x) \preceq s_2$. Conversely, suppose that $s_1 \preceq s_2$, and let $x \in F_1$; then $\operatorname{sign}(D^* x) \preceq s_1 \preceq s_2$; thus $x \in F_2$.

Assume now that these two conditions hold. Then, we have $F_1 \subset F_2 \cap \{x : J_1 \subset \operatorname{cosupp}(D^* x)\}$ since $F_1 \subset \{x : J_1 \subset \operatorname{cosupp}(D^* x)\}$ (recall Remark 2.2: $\operatorname{sign}(D^* x) \preceq s_1$ implies that $J_1 \subset \operatorname{cosupp}(D^* x)$). Conversely, let $x \in F_2 \cap \{x : J_1 \subset \operatorname{cosupp}(D^* x)\}$. Since $\operatorname{sign}(D^* x)$ and s_1 are both $\preceq s_2$, they are consistent and thus the converse is true: $J_1 \subset \operatorname{cosupp}(D^* x)$ implies that $\operatorname{sign}(D^* x) \preceq s_1$. And since $x \in F_2 \subset \partial B_1$, $x \in F_1$. The same proof holds for $\operatorname{ri}(F_1)$ and $\{x : J_1 = \operatorname{cosupp}(D^* x)\}$. For $\operatorname{dir}(F_1)$, note that $\operatorname{Ker} D^*_{J_1} \subset \operatorname{Ker} D^*_{J_2}$ (since $J_2 = \operatorname{cosupp}(s_2) \subset J_1$). Then $\operatorname{dir}(F_2) \cap \operatorname{Ker} D^*_{J_1} = (D s_2)^\perp \cap \operatorname{Ker} D^*_{J_1}$ (and $\operatorname{dir}(F_1) = (D s_1)^\perp \cap \operatorname{Ker} D^*_{J_1}$). By the proof of Lemma 2.15(ii), for $d \in \operatorname{Ker} D^*_{J_1}$, $\langle D s, d\rangle = \langle D s_1, d\rangle$ for all $s \succeq s_1$. In particular, $(D s_1)^\perp \cap \operatorname{Ker} D^*_{J_1} = (D s_2)^\perp \cap \operatorname{Ker} D^*_{J_1}$, which concludes the proof. ∎

In the spirit of [7], we consider the compact polyhedron $(\operatorname{Ker} D^*)^\perp \cap B_1$, which is isomorphic to the projection of B_1 onto the quotient of the ambient space by the lineality space $\operatorname{Ker} D^*$. Proposition 2.19 gives the following necessary and sufficient condition of extremality.

COROLLARY 2.25. *The convex polyhedron* $(\operatorname{Ker} D^*)^\perp \cap B_1$ *is compact (i.e., is a convex polytope). It admits extreme points that belong to* $(\operatorname{Ker} D^*)^\perp \cap \partial B_1$. *Given* $\bar{x} \in (\operatorname{Ker} D^*)^\perp \cap \partial B_1$, $\bar{s} = \operatorname{sign}(D^* \bar{x})$, *and* $\bar{J} = \operatorname{cosupp}(D^* \bar{x})$,

$$\bar{x} \in \operatorname{ext}((\operatorname{Ker} D^*)^\perp \cap B_1) \Leftrightarrow (\operatorname{Ker} D^*)^\perp \cap (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}} = \{0\}.$$

Remark 2.26. In [7, section 4.1.3], the authors notice that since $(\operatorname{Ker} D^*)^\perp \cap B_1$ and $\operatorname{Im} D^* \cap B_{\ell^1}$ are in bijection through D^* and its pseudo-inverse $(D^*)^+$, it holds that

$$\bar{x} \in \operatorname{ext}((\operatorname{Ker} D^*)^\perp \cap B_1) \Leftrightarrow \bar{x} = (D^*)^+ \bar{\theta} \text{ with } \bar{\theta} \in \operatorname{ext}(\operatorname{Im}\ D^* \cap B_{\ell^1}).$$

Thus we have two conditions of extremality, which are of different nature but of course equivalent, as it can be shown directly. We have already derived in Remark 2.20 that if $(\operatorname{Ker} D^*)^\perp \cap (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}} = \{0\}$, then $\bar{x} \in \operatorname{ext}((\operatorname{Ker} D^*)^\perp \cap B_1)$; it follows that $\bar{\theta} = D^* \bar{x} \in \operatorname{ext}(\operatorname{Im} D^* \cap B_{\ell^1})$ and is such that $(D^*)^+ \bar{\theta} = \bar{x}$ (recall that $(D^*)^+ D^*$ is the orthogonal projection onto $\operatorname{Im} D = (\operatorname{Ker} D^*)^\perp$). Conversely, let $\bar{\theta} \in \operatorname{ext}(\operatorname{Im} D^* \cap B_{\ell^1})$ and $\bar{x} = (D^*)^+ \bar{\theta}$ (note that $D^* \bar{x} = \bar{\theta}$ since $D^*(D^*)^+$ is the orthogonal projection onto $\operatorname{Im} D^*$). Let $d \in (\operatorname{Ker} D^*)^\perp \cap (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}}$. Note that $\operatorname{cosupp}(\bar{\theta}) = \bar{J} \subset \operatorname{cosupp}(D^* d)$. Then $\operatorname{sign}(\bar{\theta} + \varepsilon D^* d) = \bar{s}$ for $|\varepsilon|$ small, and by Lemma 2.3, $\|\bar{\theta} + \varepsilon D^* d\|_1 = \langle \bar{s}, \bar{\theta} + \varepsilon D^* d\rangle = \langle \bar{s}, \bar{\theta}\rangle + \varepsilon \langle D\bar{s}, d\rangle = \|\bar{\theta}\|_1$, i.e., $\bar{\theta} + \varepsilon D^* d \in B_{\ell^1}$ for $|\varepsilon|$ small (and $\bar{\theta} + \varepsilon D^* d \in \operatorname{Im} D^*$). By extremality of $\bar{\theta}$, $D^* d = 0$, i.e., $d \in \operatorname{Ker} D^* \cap (\operatorname{Ker} D^*)^\perp = \{0\}$, which ends the proof.

2.6. Testing the extremality. In this subsection, we aim to show that the results of subsection 2.5 can be exploited to numerically test the extremality of a point.

Our first definition formalizes the idea of feasible sign, i.e., signs which are attained by some vector in the ambient space.

Definition 2.27. *We say that a sign* $s \in \{-1,0,1\}^p$ *is* feasible *with respect to* D *if there exists* $x \in \mathbb{R}^n$ *such that* $s = \operatorname{sign}(D^* x)$.

Note that we can replace at no cost $\mathbb{R}^n$ by B_1 or ∂B_1 thanks to the homogeneity of the the ℓ^1-norm. Testing if a sign is feasible has the complexity of a linear program on n variables with p constraints.

Lemma 2.28. *Let* $c \in \mathbb{R}^n$, $s \in \{-1,0,1\}^p$, $J = \operatorname{cosupp}(x)$, $I^+ = \{i : \langle d_i, x\rangle > 0\}$, *and* $I^- = \{i : \langle d_i, x\rangle < 0\}$. *The sign* s *is feasible if, and only if, the solution set of*

$$\min_{x \in \mathbb{R}^n} \langle c, x\rangle \quad \text{subject to} \quad \begin{cases} D_J^* x & = 0 \\ D_{I^+}^* x & \geqslant +1 \\ D_{I^-}^* x & \leqslant -1 \end{cases} \tag{2.1}$$

is nonempty. Moreover, if s *is feasible, then any solution* x *of* (2.1) *is such that* $s = \operatorname{sign}(D^* x)$.

We defer to section 3 how the choice of c can leads to interesting properties. Here, we wrote the problem as a linear program to put an emphasis that existing solvers allow us to test this property. Note that finding all feasible signs is quite costly since it needs an exponential (in p) number of linear programs.

Thanks to Corollary 2.25, we have the following definition.

Definition 2.29. *We say that a sign* $s \in \{-1,0,1\}^p$ *is* pre-extreme *if it satisfies*

$$(\operatorname{Ker} D^*)^\perp \cap (D\bar{s})^\perp \cap \operatorname{Ker} D_{\bar{J}}^* = \{0\},$$

where $J = \operatorname{cosupp}(s)$, *and is* extreme *if it feasible and pre-extreme.*

Checking if a sign is pre-extreme boils down to compute the null-space of the matrix

$$B = \begin{pmatrix} U & D\bar{s} & D_{\bar{J}} \end{pmatrix}^*,$$

where U is a basis of the null space of D. In order to find the dimension of $\operatorname{Ker} B$, one can use either QR reduction or SVD (singular value decomposition). Here, we used an SVD approach.

We now illustrate these definitions in low dimension (it is known that the study of the number of faces is a very difficult task in general [20, 4]), when D corresponds to the incidence matrix of a complete graph on $n = 4$ vertices and $p = 6$ edges. Among $3^p = 729$ possible signs, only 75 are feasible, and among them 14 are extreme. We report in Figure 2.1 the pattern of such signs up to centrosymmetry of the unit ball; i.e., we only show 7 of the 14 extreme signs.

3. The solution set of sparse analysis regularization. This section is an application of the previous one towards the solution set of (1.1). Remark that a similar analysis can be performed (in a less challenging way) for the noiseless problem (1.2). In the first subsection 3.1, we show that the solution set can be seen as a particular subpolyhedron of the unit ball. Using this result, we derive several structural results in subsection 3.2 on the solution set thanks to section 2. Finally, we show that arbitrary subpolyhedra of the unit ball can be seen as solution set of (1.1) with a specific choice of parameters.

3.1. The solution set as a subpolyhedron of the unit ball. This section studies the structure of the solution set of (1.1) that we denote by X:

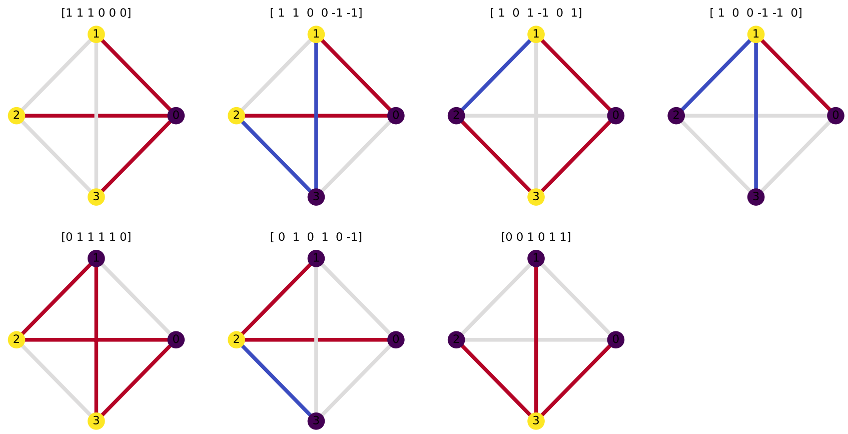

FIG. 2.1. *Extreme signs of a complete graph on $n=4$ vertices. Red edges correspond to positive sign, blue edges to negative sign, and gray to* 0.

$$X = \operatorname*{argmin}_{x\in\mathbb{R}^n} \mathcal{L}(x) \overset{\text{def.}}{=} \frac{1}{2}\|y-\Phi x\|_2^2 + \lambda\|D^* x\|_1,$$

where $y \in \mathbb{R}^q$, $\Phi\colon \mathbb{R}^n \to \mathbb{R}^q$, $D\colon \mathbb{R}^p \to \mathbb{R}^n$ is linear and $\lambda > 0$.

THEOREM 3.1. *The solution set of* (1.1) *is a nonempty convex polyhedron of the form*

$$X = \mathcal{A} \cap B_r$$

with $r \geq 0$ and $\mathcal{A}$ an affine subspace such that $\emptyset \neq \mathcal{A}\cap B_r \subset \partial B_r$ and $\operatorname{dir}(\mathcal{A}) = \operatorname{Ker}\Phi$. *Namely, if $x \in X$, then $r = \|D^*x\|_1$ and $\mathcal{A} = x + \operatorname{Ker}\Phi$.*

Proof. It is easy to see that the objective function $x \mapsto \mathcal{L}(x)$ is a nonnegative, convex, closed continuous function with full domain (in particular proper). However, it is not coercive; hence the existence of minimizers and compactness of the solution set are not straightforward. Following [22, Chapter 8], the recession cone $R_\mathcal{L}$ of $\mathcal{L}$ is given by

$$R_\mathcal{L} \overset{\text{def.}}{=} \{z \in \mathbb{R}^n \,:\, \mathcal{L}_\infty(z) \leqslant 0\},$$

where $\mathcal{L}_\infty$ is the recession function of $\mathcal{L}$ given by

$$\mathcal{L}_\infty(z) \overset{\text{def.}}{=} \lim_{t\to+\infty} \frac{\mathcal{L}(tz)}{t} \in \mathbb{R}\cup\{+\infty\}.$$

It is clear that $R_\mathcal{L}$ is nonnegative; hence the recession cone $R_\mathcal{L}$ is given by $R_\mathcal{L} = \{z \in \mathbb{R}^n \;:\; \mathcal{L}_\infty(z) = 0\}$. The lineality space $L_\mathcal{L}$ is the subspace of $\mathbb{R}^n$ formed by elements d such that $d \in R_\mathcal{L}$ and $-d \in R_\mathcal{L}$, i.e, $L_\mathcal{L} = R_\mathcal{L} \cap (-R_\mathcal{L})$. The following lemma characterizes the structure of $R_\mathcal{L}$ and $L_\mathcal{L}$.

LEMMA 3.2. *The recession cone $R_\mathcal{L}$ and the lineality space $L_\mathcal{L}$ of $\mathcal{L}$ are given by*

$$R_\mathcal{L} = L_\mathcal{L} = \operatorname{Ker} D^* \cap \operatorname{Ker}\Phi.$$

Proof. Let $t > 0$ and $z \in \mathbb{R}^n$. We have

$$\frac{1}{t}\mathcal{L}(tz) = \frac{1}{2t}\|y - \Phi t z\|_2^2 + \lambda\|D^* z\|_1$$

$$= \frac{1}{2t}\|y\|_2^2 - \langle y,\, \Phi z\rangle + t\|\Phi z\|_2^2 + \lambda\|D^* z\|_1.$$

Hence,

$$\mathcal{L}_\infty(z) = -\langle y,\, \Phi z\rangle + \lambda\|D^* z\|_1 + \iota_{\operatorname{Ker}\Phi}(z).$$

In particular, $\mathcal{L}_\infty(z) = 0$ if, and only if, $z \in \operatorname{Ker} D^* \cap \operatorname{Ker}\Phi$. Since $\operatorname{Ker} D^* \cap \operatorname{Ker}\Phi$ is a subspace, we have $R_{\mathcal{L}} = L_{\mathcal{L}}$. ∎

The polyhedral structure of X, as we will see, relies on the following lemma.

Lemma 3.3. *Let $C \subset \mathbb{R}^n$ be a nonempty convex set. Then $\mathcal{L}$ is constant on C if and only if Φ and $\|D^* \cdot\|_1$ are constant on C.*

Proof. Assume that $\mathcal{L}$ is constant on C, and let $x_0 \in C$. Suppose that there exists $x_1 \in C$ such that $\Phi x_1 \neq \Phi x_0$, and let $x = \frac{x_0+x_1}{2}$ (note that $x \in C$). Then by strict convexity of $u \mapsto \|y-u\|_2^2$,

$$\begin{aligned}\big\|y - \Phi x\big\|_2^2 &= \left\|y - \left(\frac{1}{2}\Phi x_0 + \frac{1}{2}\Phi x_1\right)\right\|_2^2 \\ &< \frac{1}{2}\big\|y - \Phi x_0\big\|_2^2 + \frac{1}{2}\big\|y - \Phi x_1\big\|_2^2.\end{aligned}$$

Together with the convexity inequality of the ℓ^1 norm

$$\big\|D^* x\big\|_1 \leq \frac{1}{2}\big\|D^* x_0\big\|_1 + \frac{1}{2}\big\|D^* x_1\big\|_1,$$

we get that $\mathcal{L}(x) < \frac{1}{2}\mathcal{L}(x_0) + \frac{1}{2}\mathcal{L}(x_1)$, which is in contradiction with $\mathcal{L}$ constant on C. Then Φ is constant on C and thus $\|D^* \cdot\|_1$ too. The converse is straightforward. ∎

We add the following lemma that gives locally the directions where $\|D^* \cdot\|_1$ is constant.

Lemma 3.4. *Let $\mathcal{A}$ be an affine subspace and $\bar{x} \in \mathcal{A}$. Then $\|D^* \cdot\|_1$ is constant in a neighborhood of $\bar{x}$ in $\mathcal{A}$ if and only if*

$$\operatorname{dir}(\mathcal{A}) \subset (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}},$$

where $\bar{s} = \operatorname{sign}(D^*\bar{x})$ and $\bar{J} = \operatorname{cosupp}(D^*\bar{x})$.

Proof. By definition, $\|D^* \cdot\|_1$ is constant in a neighborhood of $\bar{x}$ in $\mathcal{A}$ if and only if there exists $\varepsilon > 0$ such that $B(\bar{x},\varepsilon) \cap \mathcal{A} \subset \partial B_r$ with $r = \|D^*\bar{x}\|_1$. We denote by $C = B(\bar{x},\varepsilon) \cap \mathcal{A}$ (note that $\bar{x} \in \operatorname{ri}(C)$). By Proposition 2.9, $C \subset \partial B_r$ if and only if $C \subset \bar{F}$ with $\bar{F} = B_r \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = 1\}$ if $r > 0$ and $\bar{F} = \operatorname{Ker} D^*$ if $r = 0$. Since $\bar{x} \in \operatorname{ri}(\bar{F})$, $C \subset \bar{F}$ if and only if $\operatorname{dir}(C) \subset \operatorname{dir}(\bar{F})$. The result follows by noticing that $\operatorname{dir}(C) = \operatorname{dir}(\mathcal{A})$ (e.g., by Lemma 2.14) and $\operatorname{dir}(\bar{F}) = (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}}$ (by Lemma 2.8 in the case $r > 0$, which is obvious in the case $r = 0$). ∎

Together, the previous two lemmas give the following.

Corollary 3.5. *Let $\mathcal{A}$ be an affine subspace and $\bar{x} \in \mathcal{A}$. Then $\mathcal{L}$ is constant in a neighborhood of $\bar{x}$ in $\mathcal{A}$ if and only if*

$$\operatorname{dir}(\mathcal{A}) \subset \operatorname{Ker}\Phi \cap (D\bar{s})^\perp \cap \operatorname{Ker} D^*_{\bar{J}},$$

where $\bar{s} = \operatorname{sign}(D^*\bar{x})$ and $\bar{J} = \operatorname{cosupp}(D^*\bar{x})$.

We now go back to the proof of Theorem 3.1. By Lemma 3.2, the recession cone and the lineality space of $\mathcal{L}$ coincide. Then by [22, Theorem 27.1(a–b)], the solution set X is nonempty. Since $\mathcal{L}$ is convex (and closed), the solution set is also convex (and closed). Moreover, $\mathcal{L}$ is constant on X. Then by Lemma 3.3, Φ and $\|D^* \cdot\|_1$ are constant on X, i.e.,

$$X \subset (x + \operatorname{Ker}\Phi) \cap \partial B_r$$

with $x \in X$ and $r = \|D^* x\|_1$. But since $\mathcal{L}(x)$ is the minimum of $\mathcal{L}$,

$$(x + \operatorname{Ker}\Phi) \cap B_r \subset X.$$

It follows that

$$X = (x + \operatorname{Ker}\Phi) \cap B_r$$

with $(x + \operatorname{Ker}\Phi) \cap B_r \subset \partial B_r$, as it was to be proved.

3.2. Consequence results on the solution set. In this section, we apply the results on subpolyhedra of the unit ball (subsection 2.4) to the solution set X of (1.1).

PROPOSITION 3.6. *Let $\bar{x} \in \operatorname{ri}(X)$, $\bar{s} = \operatorname{sign}(D^*\bar{x})$, and $\bar{F} = B_r \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = r\}$ with $r = \|D^*\bar{x}\|_1$. Then*

$$X = (\bar{x} + \operatorname{Ker}\Phi) \cap \bar{F}.$$

It follows that

$$\begin{aligned} X &= (\bar{x} + \operatorname{Ker}\Phi) \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq \bar{s}\}, \\ \operatorname{ri}(X) &= (\bar{x} + \operatorname{Ker}\Phi) \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) = \bar{s}\}, \\ \operatorname{dir}(X) &= \operatorname{Ker}\Phi \cap \operatorname{Ker} D^*_{\bar{J}} \ \textit{(where } \bar{J} = \operatorname{cosupp}(\bar{s})\textit{)}. \end{aligned}$$

Moreover, the faces of X are exactly the sets of the form $\{x \in X : J \subset \operatorname{cosupp}(D^*x)\}$ with $\bar{J} \subset J$; their relative interior is given by $\{x \in X : J = \operatorname{cosupp}(D^*x)\}$ and their direction by $\operatorname{Ker}\Phi \cap \operatorname{Ker} D^*_J$.

Proof. First note that the results are trivial in the case $r = 0$, as $\bar{s} = 0$ and $\bar{F} = B_r = \operatorname{Ker} D^*$. Therefore we consider the case $r > 0$. The first statement follows from Theorem 3.1 and Proposition 2.17, and the second one from Theorem 2.16(iii). For the last statement, note that G is a face of X if and only if $G = (\bar{x} + \operatorname{Ker}\Phi) \cap F$ with F a face of B_r such that $F \subset \bar{F}$. Indeed, the direct implication holds for $G = \emptyset$ with $F = \emptyset$, for $G = X$ with $F = \bar{F}$, and for G exposed in $\operatorname{aff}(X)$ by Proposition 2.18. Conversely, let $F = \emptyset$ or $B_r \cap \mathcal{H}$ (with $\mathcal{H}$ a supporting hyperplane) be a face of B_r. Then $(\bar{x} + \operatorname{Ker}\Phi) \cap F = \emptyset$ or $X \cap \mathcal{H}$ is a face of X. The conclusion follows from Proposition 2.24. ∎

Thanks to Proposition 3.6, we can draw several conclusions. In particular the role of $\bar{s}$ and $\bar{J}$ allows us to derive properties of the solution set.

The sign $\bar{s}$ is shared by all the interior solutions of (1.1), which are also maximal solutions ($\bar{s} = \max_{x \in X} \operatorname{sign}(D^*x)$). Such a solution can be obtained numerically by the algorithm described in [2]. Future work should include an analysis of the behavior of more common algorithms such as first-order proximal methods.

The knowledge of $\bar{s}$ (or of $\bar{J}$ which is the minimal cosupport) gives the dimension of the solution set (1.1) as $\dim(X) = \dim(\operatorname{Ker}\Phi \cap \operatorname{Ker} D_{\bar{J}}^*)$ (up to determining the dimension of the null space of the matrix $A = \begin{pmatrix}\Phi^* & D_{\bar{J}}\end{pmatrix}^*$, which again can be done by QR reduction or SVD). For instance, taking the example of [2, section 6.3],

$$D^* = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 2 & 1 & 1 \end{pmatrix}, \quad \Phi = \begin{pmatrix} 1 & 1 & 1 \\ 3 & 1 & 1 \\ \sqrt{2} & 0 & 0 \end{pmatrix}, \quad y = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad \lambda = \frac{1}{2}, \tag{3.1}$$

we can prove that $X = \operatorname{conv}(0 \quad \frac{1}{2} \quad 0)^*, (0 \quad 0 \quad \frac{1}{2})^*$. But running the interior-point method described in [2] leads to the specific solution (up to numerical error) $\bar{x} = (0 \quad \frac{1}{4} \quad \frac{1}{4})^*$. In this case, the matrix A reduces to $A = \Phi$ since $\bar{J} = \emptyset$. Thus, $\dim(X) = \dim(\operatorname{Ker}\Phi) = 1$.

In the previous formula, $\dim(X)$ is decreasing w.r.t. $\bar{J}$; it somehow quantifies the tautology according to which the less sparse solution, the fewer solutions.

Together with Lemma 2.6, the previous proposition gives the following half-space representation of the solution set X of (1.1) (one could of course give similar representations of its faces).

Corollary 3.7. *Let $\bar{x} \in \operatorname{ri}(X)$, $\bar{s} = \operatorname{sign}(D^*\bar{x})$, $\bar{J} = \operatorname{cosupp}(\bar{s})$ and $\bar{I} = \operatorname{supp}(\bar{s})$. Then*

$$X = \{x \in \mathbb{R}^n : \Phi x = \Phi\bar{x}, D_{\bar{J}}^* x = 0, \operatorname{diag}(\bar{s}_{\bar{I}}) D_{\bar{I}}^* x \geq 0\}.$$

This result can be used numerically. Indeed, it provides a linear characterization of the solution set up to the knowledge of a maximal solution. In the same spirit of [32], we can derive bounds on the coefficients (both in the signal domain or in the dictionary domain). For instance, finding the biggest i-coefficient boils down to solve the linear program

$$\max_{x \in X} \langle x, e_i \rangle,$$

where e_i the is the ith canonical vector. Thus, we can describe in a similar fashion which component are dispensable following the vocabulary introduced in [32].

We can also apply Theorem 2.16 to an arbitrary solution (not necessarily interior). This result is useful when obtaining a solution computed from any algorithm without guarantees on its maximality.

Proposition 3.8. *Let $x \in X$, $s = \operatorname{sign}(D^*x)$, $F = B_r \cap \{x \in \mathbb{R}^n : \langle Ds, x\rangle = r\}$ with $r = \|D^*x\|_1$. Then*

$$G = (x + \operatorname{Ker}\Phi) \cap F$$

is the smallest face of X such that $x \in G$ and the unique face of X such that $x \in \operatorname{ri}(X)$. It satisfies

$$G = (x + \operatorname{Ker}\Phi) \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) \preceq s\},$$
$$\operatorname{ri}(G) = (x + \operatorname{Ker}\Phi) \cap \{x \in \mathbb{R}^n : \operatorname{sign}(D^*x) = s\},$$
$$\operatorname{dir}(G) = \operatorname{Ker}\Phi \cap \operatorname{Ker} D_J^* \textit{ (where } J = \operatorname{cosupp}(s)\textit{)}.$$

Note in particular that $\dim(\operatorname{Ker}\Phi \cap \operatorname{Ker} D_J^*)$ is always the dimension of a subset of solutions. When $D = \operatorname{Id}$, then $\operatorname{Ker}\Phi \cap \operatorname{Ker} D_J^* = \operatorname{Ker}\Phi_I$ (with $I = [n] \setminus J$ the support of x) the rank deficiency of Φ_I (the difference between the size of the support I and the rank of Φ_I) is a lower bound of the dimension of the solution set. See subsection 3.6 for an illustration on a real data set.

We end this section with a characterization of the compactness of X and of its extreme points, as well as a sufficient condition for uniqueness knowing a solution.

PROPOSITION 3.9. *The solution set X of* (1.1) *admits extreme points if and only if it is compact (i.e., is a convex polytope), if and only if*

$$\operatorname{Ker}\Phi \cap \operatorname{Ker} D^* = \{0\}.$$

A solution $x \in X$ is an extreme point (i.e., $x \in \operatorname{ext}(X)$) if and only if, denoting by $J = \operatorname{cosupp}(D^* x)$,

$$\operatorname{Ker}\Phi \cap \operatorname{Ker} D_J^* = \{0\}.$$

Proof. Recall that $X = (\bar{x} + \operatorname{Ker}\Phi) \cap B_r$ by Theorem 3.1. The result is trivial in the case $r = 0$ and is the transcription of Proposition 2.19 in the case $r > 0$. ∎

COROLLARY 3.10. *Let $x \in X$ and $J = \operatorname{cosupp}(D^* x)$. If x is the unique solution of* (1.1) *(i.e., $X = \{x\}$), then* $\operatorname{Ker}\Phi \cap \operatorname{Ker} D_J^* = \{0\}$.

The condition $\operatorname{Ker}\Phi \cap \operatorname{Ker} D^* = \{0\}$ is equivalent to the recession cone $R_{\mathcal{L}}$ being reduced to $\{0\}$, which is known to be equivalent to the compactness of the solution set X of (1.1); see, e.g., [22, Theorem 27.1(d)]. Note that this is the condition denoted by (H_0) and assumed to hold all throughout [36] or [35]. This condition can be specified to our examples:

- When D is the identity $D = \operatorname{Id}$, it is automatically satisfied;
- When $D = D_{TV}$, this condition is satisfied as soon as Φ does not cancel on constant vectors;
- When D is the incidence matrix of a graph, observe that this condition reduces to the fact that Φ should not be constant on the set of constant vectors in each connected component.

The condition $\operatorname{Ker}\Phi \cap \operatorname{Ker} D_J^* = \{0\}$ is the one denoted by (H_J) and required in [35] at a given solution in order to undertake a sensitivity analysis. It turns out from the present study that such solutions are precisely the extreme points of the solution set. In [35, section A.3], an iterative procedure is proposed to construct such an extreme point. Alternatively, if one has the knowledge that the maximal solution is quite sparse, then an exhaustive test can be performed in a similar fashion than subsection 2.6.

Going back to the setting proposed in (3.1), we observe that thanks to Proposition 3.9, we know that X is compact since $\operatorname{Ker}\Phi$ and $\operatorname{Ker} D^*$ intersect trivially, and we can obtain the extreme points by observing that there is three feasible signs $(1,0,1)$, $(0,1,1)$, and $(1,1,1)$. Only the first two leads to a cosupport J such that $\operatorname{Ker}\Phi$ and $\operatorname{Ker} D_J^*$ intersect trivially. Using Corollary 3.7, one can use any linear solver from these two signs to obtain associated two extreme points of the solution set, i.e., $\operatorname{ext}(X) = \{(0 \quad \frac{1}{2} \quad 0)^*, (0 \quad 0 \quad \frac{1}{2})^*\}$.

3.3. Discussion of uniqueness conditions. We can derive from Proposition 3.6 a necessary and sufficient condition for uniqueness. Indeed (1.1) admits a unique solution if and only if $\operatorname{dir} X = \{0\}$. Therefore if one knows the minimal cosupport $\bar{J}$ (that is the cosupport of a maximal solution $\bar{x}$), then (1.1) admits a unique solution if and only $\operatorname{Ker}\Phi \cap \operatorname{Ker} D_{\bar{J}}^* = \{0\}$. Recall that a maximal solution can be obtained numerically by the algorithm described in [2], so together with QR reduction or SVD, uniqueness can be checked numerically.

In the case of the Lasso problem ($D = \mathrm{Id}$), this necessary and sufficient condition becomes $\operatorname{Ker}\Phi_{\bar{I}} = \{0\}$ with $\bar{I}$ the maximal support. It is slightly weaker than the first sufficient condition for uniqueness in the seminal paper of Tibshirani [32, Lemma 2]. Indeed, the latter is $\operatorname{Ker}\Phi_{\mathcal{E}} = \{0\}$ with $\mathcal{E}$ the so-called equicorrelation set, which always contains the maximal support $\bar{I}$ and coincides with it for almost every y [32, Lemma 13]. In this same paper, Tibshirani proves that his first sufficient condition is satisfied if the columns of Φ are in *general position* [32, Lemma 3], which is the case with probability one if their entries are drawn from a continuous distribution [32, Lemma 4]. Note that we make no such assumption in our paper. The fact that nonuniqueness may arise when Φ has discrete entries is confirmed by Ewald and Schneider [12, Theorem 14]: there exists y for which the Lasso problem has a nonunique solution if and only if $\operatorname{Im}(\Phi^*)$ intersects a face of $[-1,1]^n$ of dimension strictly smaller than the dimension of $\operatorname{Ker}(\Phi)$; it is in particular the case when Φ has a row with only ± 1 entries and a nontrivial null space.

In the general case, the derivation of uniqueness conditions in the paper of Ali and Tibshirani [1] is complicated by the fact that there is no (unique) equicorrelation set but several *boundary sets* $\mathcal{B}$ for which $\operatorname{Ker}\Phi \cap \operatorname{Ker} D^*_{[n]\setminus\mathcal{B}} = \{0\}$ needs to be satisfied. The authors introduced a notion of D^*-*general position* which, together with the condition $\operatorname{Ker}\Phi \cap \operatorname{Ker} D^* = \{0\}$ (equivalent to the compactness of the solution set by Proposition 3.6), implies uniqueness for almost every y [1, Lemma 6]. These conditions are satisfied with probability one when the entries of Φ are drawn from a continuous distribution and $n \leq q$ or $n > q$ and $\dim(\operatorname{Ker} D^*) \leq q$ [1, Lemmas 7 and 8] but are not assumed in our work. Finally, the necessary and sufficient condition of Ewald and Schneider for the uniqueness of the Lasso minimizer above has been generalized by Schneider and Tardivel to regularizations by polyhedral norms [24] and then by polyhedral gauges [29]. The latter framework includes the generalized Lasso for which their condition is that there exists y such that (1.1) has a nonunique solution if and only if $\operatorname{Im}(\Phi^*)$ intersects a face of $D[-1,1]^n$ (the image of the hypercube by D, which is a polyhedron in $\mathbb{R}^q$) of dimension strictly smaller than the dimension of $\operatorname{Ker}(\Phi)$.

3.4. Arbitrary subpolyhedra of the unit ball as solution sets. We have the following converse of Theorem 3.1.

THEOREM 3.11. *Let $r \geq 0$ and $\mathcal{A}$ be an affine subspace such that $\emptyset \neq \mathcal{A} \cap B_r \subset \partial B_r$. Then there exist Φ, y, and $\lambda > 0$ such that the solution set of* (1.1) *is $X = \mathcal{A} \cap B_r$ and $\operatorname{Ker}\Phi = \operatorname{dir}(\mathcal{A})$.*

Proof. We first consider the case $r > 0$. Let $\bar{x} \in \mathcal{A} \cap \partial B_r$ and $\bar{s} = \operatorname{sign}(D^*\bar{x})$. We define Φ, y and λ as follows: we consider $a_1, \ldots, a_m$ a basis of $\operatorname{dir}(\mathcal{A})^\perp$ (with m the dimension of this subspace) and set $\Phi = (a_1 | \cdots | a_m)^*$; then $\operatorname{Im}(\Phi^*) = \operatorname{dir}(A)^\perp$ and $\operatorname{Ker}(\Phi) = \operatorname{dir}(\mathcal{A})$. It follows from Lemma 2.15(ii) that $\operatorname{Im}(\Phi^*) \cap (\sum_{s \succeq \bar{s}} \mathbb{R}_+ Ds) \neq \{0\}$. Let $\beta \in \mathbb{R}^q$ and $\alpha_s \geq 0$ for all $s \succeq \bar{s}$ be such that

$$\Phi^*\beta = \sum_{s \succeq \bar{s}} \alpha_s Ds \neq 0;$$

we can assume (by normalizing) that $\sum_{s \succeq \bar{s}} \alpha_s = 1$. We define $u = \sum_{s \succeq \bar{s}} \alpha_s s$, so that $Du = \Phi^*\beta$. Note also that $u_{\bar{I}} = \bar{s}_{\bar{I}}$ and $\|u_{\bar{J}}\|_\infty \leq 1$ (with $\bar{I} = \operatorname{supp}(D^*\bar{x})$ and $\bar{J} = \operatorname{cosupp}(D^*\bar{x})$). We now fix any $\lambda > 0$ and set $y = \Phi\bar{x} + \lambda\beta$. We denote as always by X the solution set of (1.1). By construction, we have

$$\Phi^*(\Phi\bar{x} - y) + \lambda Du = 0.$$

It implies that $0 \in \partial\mathcal{L}(\bar{x})$ since $\partial(\|D^* \cdot\|_1)(\bar{x}) = D\partial(\|\cdot\|_1)(D^*\bar{x})$ and $u \in \partial(\|\cdot\|_1)(D^*\bar{x})$ (see, e.g., [14] or [35]), and thus $\bar{x} \in X$. It follows from Theorem 3.1 that $X = (\bar{x} + \operatorname{Ker}\Phi) \cap B_{\|D^*\bar{x}\|_1}$. But $\bar{x} + \operatorname{Ker}\Phi = \mathcal{A}$ since $\bar{x} \in \mathcal{A}$ and $\operatorname{Ker}\Phi = \operatorname{dir}(\mathcal{A})$, and $\|D^*\bar{x}\|_1 = r$ since $\bar{x} \in \partial B_r$, which concludes the proof of the case $r > 0$.

We now treat the case $r = 0$ for which $B_r = \partial B_r = \operatorname{Ker} D^*$. We define Φ as in the previous case, $\lambda > 0$ arbitrarily, and $y = \Phi\bar{x}$ for some $\bar{x} \in \mathcal{A} \cap \operatorname{Ker} D^*$. Then again, $\Phi^*(\Phi\bar{x} - y) + \lambda D u = 0$, here with $u = 0$ ($\|u\|_\infty \leq 1$). It follows that $\bar{x} \in X$, and then that $X = \mathcal{A} \cap B_r$ as before. □

If we relax the condition $\operatorname{Ker}\Phi = \operatorname{dir}(\mathcal{A})$, we can get rid of the assumption $\emptyset \neq \mathcal{A} \cap B_r \subset \partial B_r$ and at the same time choose the exposed face F so that $\mathcal{A} \cap F$ is a solution set (thus of arbitrary dimension). This is the object of the next proposition.

PROPOSITION 3.12. *Let $r > 0$, F be an exposed face of B_r and $\mathcal{A}$ be an affine subspace intersecting F. Then there exist Φ, y, and $\lambda > 0$ such that the solution set of* (1.1) *is $X = \mathcal{A} \cap F$ (and $\operatorname{Ker}\Phi \subset \operatorname{dir}(\mathcal{A})$).*

Proof. Let $\bar{s} = \max_{x \in F} \operatorname{sign}(D^*x)$, so that $F = B_r \cap \{x \in \mathbb{R}^n : \langle D\bar{s}, x\rangle = r\}$ by Proposition 2.21. Let $\Phi = (D\bar{s}|a_1|\cdots|a_m)^*$ with $a_1, \ldots, a_m$ a basis of $\operatorname{dir}(\mathcal{A})^\perp$, so that $\ker\Phi = (D\bar{s})^\perp \cap \operatorname{dir}(\mathcal{A})$; let $\lambda > 0$, $y = \Phi x + \lambda e_1$ for some $x \in \mathcal{A} \cap F$, and X be the associated solution set. Then $\Phi^*(\Phi x - y) + \lambda D u = 0$ with $u = \bar{s}$. Since $\operatorname{sign}(D^*x) \preceq \bar{s}$, $u_I = \operatorname{sign}(D^*x)_I$ and $\|u_J\|_\infty \leq 1$ (with $I = \operatorname{supp}(D^*x)$ and $J = \operatorname{cosupp}(D^*x)$). Then $x \in X$ and $X = (x + \operatorname{Ker}\Phi) \cap B_r = (x + \operatorname{dir}(\mathcal{A})) \cap (x + (D\bar{s})^\perp) \cap B_r = \mathcal{A} \cap F$. □

We now illustrate Proposition 3.12 on nonperiodic total variation on 3 points in order to give an intuition of the geometric construction; see Figure 3.1. Let D be a forward discrete difference operator on 3 points, i.e.,

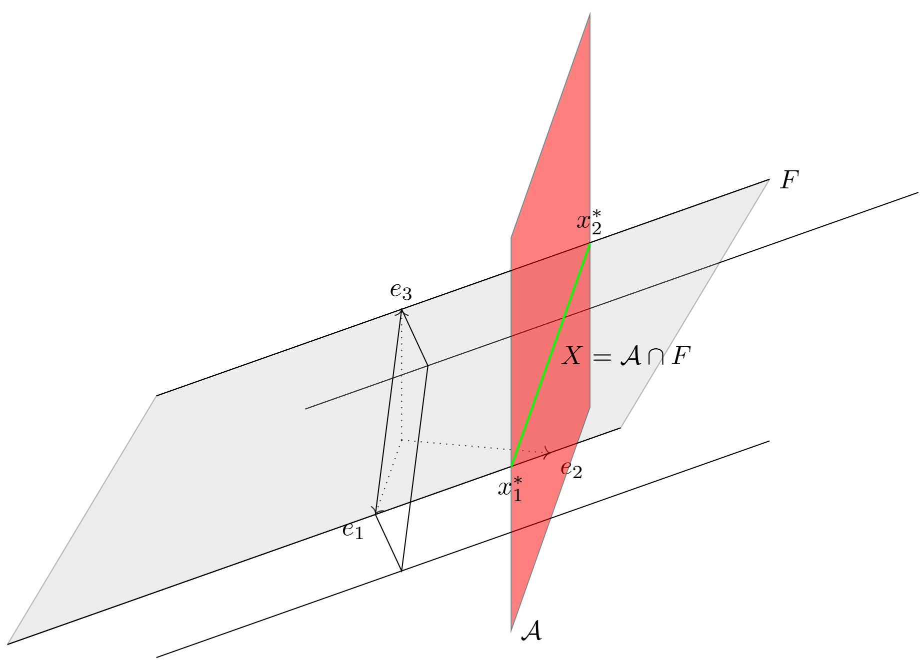

FIG. 3.1. *Construction of a solution set for the total variation regularization on 3 points and the associated Φ and y.*

$$D^* = \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}.$$

Consider the facet F (in gray on the figure) determined by the sign $\bar{s} = (-1, 1)$ and an (affine) hyperplane $\mathcal{A}$ (in red on the figure) with normal vector $(0, 1, 0)$ and origin $(1, 1, 1)$. The intersection (in green on the figure) of F and $\mathcal{A}$ is then the segment defined by $x_1^* = (1, 1, 2)$ and $x_2^* = (2, 1, 1)$. The proof of Proposition 3.12 gives us how to design a setting such that the solution set X is exactly $X = \operatorname{conv}\{x_1^*, x_2^*\}$: let $\lambda = 1$,

$$\Phi = \begin{pmatrix} 1 & -2 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

Checking first-order condition of this setting is tedious, but doable, and leads to $X = \operatorname{conv}\{x_1^*, x_2^*\}$.

3.5. Illustration of the main results for 1D total variation. We now provide a full illustration of our results in higher dimension for the popular regularization that is 1D total variation. Note that D^* is a matrix of rank $n-1$ whose null space is formed by constant vectors $\operatorname{Ker} D^* = \mathbb{R}e$, where $e = \begin{pmatrix} 1 & \ldots & 1 \end{pmatrix}^*$.

Let $2 \leqslant t_1 < t_2 < t_3 < t_4 \leqslant n-1$ and consider the reference signal $\bar{x} \in \mathbb{R}^n$ and its associated sign $\bar{s} = \operatorname{sign}(D^* \bar{x}) \in \{-1, 0, +1\}^{n-1}$ defined by

$$\bar{x}_v = \frac{1}{6} \times \begin{cases} -1 & \text{if } 1 \leqslant v \leqslant t_1, \\ 1 & \text{if } t_1 < v \leqslant t_2, \\ 0 & \text{if } t_2 < v \leqslant t_3, \\ 1 & \text{if } t_3 < v \leqslant t_4, \\ -1 & \text{if } t_4 < v \leqslant n, \end{cases} \quad \text{and} \quad \bar{s}_e = \begin{cases} 1 & \text{if } e = t_1 \text{ or } e = t_3, \\ -1 & \text{if } e = t_2 \text{ or } e = t_4, \\ 0, & \text{otherwise.} \end{cases}$$

Our objective is to build a problem (i.e, find Φ, λ and y) such that

1. $\bar{x}$ is a maximal solution (i.e., $\bar{x}$ lives in the relative interior of X).
2. Every solutions share a common jump at t_1.
3. The affine hull of X is of dimension 1.

The first step is to define a candidate extreme point of X which should be compatible with the sign $\bar{s}$ of $\bar{x}$ according to Corollary 2.12. Such vector $\hat{x}$ and sign $\hat{s}$ can be chosen as

$$\hat{x}_v = \frac{1}{6} \times \begin{cases} -1 & \text{if } 1 \leqslant v \leqslant t_1, \\ 1 & \text{if } t_1 < v \leqslant t_2, \\ 1 & \text{if } t_2 < v \leqslant t_3, \\ 1 & \text{if } t_3 < v \leqslant t_4, \\ -3 & \text{if } t_4 < v \leqslant n, \end{cases} \quad \text{and} \quad \hat{s}_e = \begin{cases} 1 & \text{if } e = t_1, \\ -1 & \text{if } e = t_4, \\ 0, & \text{otherwise.} \end{cases}$$

We are now following the proof of Proposition 3.12 to construct our sparse analysis problem. We consider the direction $d = z - x$ and build a basis $a_1, \ldots, a_{n-1}$ of $d^\perp$. This can be done either by hand or using an SVD. We then consider an arbitrary $\lambda > 0$, $\Phi = (D\bar{s}|a_1|\cdots|a_m)^*$ and $y = \Phi\bar{x} + \lambda e_1$. By construction, $\bar{x}$ and $\hat{x}$ are solutions of (1.1), and $\bar{x}$ is a maximal solution. Since Φ has rank $n-1$, the affine hull solution set X as at most dimension 1 according to Proposition 3.6. It is indeed its dimension using an SVD of $\begin{pmatrix} \Phi \\ D_{\bar{J}}^* \end{pmatrix}$, where $\bar{J} = \operatorname{supp}(D^* \bar{x})$. To fully describe it, we have to find its two extreme points.

In order to do it, we are going to use Corollary 3.7. Let $\bar{J} = \operatorname{cosupp}(\bar{s})$, $\bar{I} = \bar{J}^c$,

$$A_{\text{eq}} = \begin{pmatrix} \Phi \\ D^*_{\bar{J}} \end{pmatrix}, \quad b_{\text{eq}} = \begin{pmatrix} \Phi\bar{x} \\ 0 \end{pmatrix}, \quad \text{and} \quad A_{\text{ineq}} = -\operatorname{diag}(\bar{s}_{\bar{I}}) D^*_{\bar{I}}.$$

Consider now the linear programs for $1 \leqslant i \leqslant n-1$:

$$(\text{mLP}_i) \qquad \operatorname*{argmin}_{x \in \mathbb{R}^n} \langle x, d_i \rangle \quad \text{subject to} \quad A_{\text{eq}} x = b_{\text{eq}} \text{ and } A_{\text{ineq}} x \leqslant 0,$$

and

$$(\text{MLP}_i) \qquad \operatorname*{argmax}_{x \in \mathbb{R}^n} \langle x, d_i \rangle \quad \text{subject to} \quad A_{\text{eq}} x = b_{\text{eq}} \text{ and } A_{\text{ineq}} x \leqslant 0.$$

For a given $1 \leqslant i \leqslant n-1$, three cases may occur

1. The value of a minimizer of (mLP_i) and a maximizer (MLP_i) is 0. It means that *every* solution $x \in X$ is zero at the index i: for every $x \in X$, $\langle x, d_i \rangle = 0$.

2a. The value of a minimizer m of (mLP_i) is negative $m < 0$ and a maximizer (MLP_i) is 0. It means that *some* solution $x \in X$ is zero at the index i; i.e., there exists $x \in X$, $\langle x, d_i \rangle = 0$. We may call the index i a dispensable index to be consistent with the work [32].

2b. A symmetric situation is when the value of a minimizer of (mLP_i) is 0 and the value M of a maximizer (MLP_i) is positive. It means also that *some* solution $x \in X$ is zero at the index i.

3a. The value of a minimizer m of (mLP_i) is negative $m < 0$, and the value M of a maximizer (MLP_i) is also negative $M < 0$. It means that *no* solution $x \in X$ is sparse at the index i; i.e., for every $x \in X$, $m \leqslant \langle x, d_i \rangle \leqslant M \neq 0$. The index is called indispensable.

3b. The value of a minimizer m of (mLP_i) is positive $m > 0$ and the value M of a maximizer (MLP_i) is also positive $M > 0$. Then i is indispensable.

Note that, according to Corollary 2.12, the signs of every solution must be consistent; hence it is impossible to have the situation where the value of a minimizer m of (mLP_i) is strictly negative $m < 0$ and the value M of a maximizer (MLP_i) is strictly negative $M > 0$.

Beyond the value of (mLP_i) and (MLP_i), the actual solution of the linear program is itself a solution of (1.1). Thus, to find a second candidate to be the extreme point of X, it is sufficient to run (mLP_i) and (MLP_i) for each $1 \leqslant i \leqslant n-1$ and consider their nonzero value solutions. Doing so (using for instance `scipy.optimize.linprog` [37]) let us consider

$$\tilde{x}_v = \frac{1}{6} \times \begin{cases} -1 & \text{if } 1 \leqslant v \leqslant t_1, \\ 1 & \text{if } t_1 < v \leqslant t_2, \\ -1 & \text{if } t_2 < v \leqslant t_3, \\ 1 & \text{if } t_3 < v \leqslant t_4, \\ 1 & \text{if } t_4 < v \leqslant n, \end{cases} \quad \text{and} \quad \tilde{s}_e = \begin{cases} 1 & \text{if } e = t_1 \text{ or } e = t_3, \\ -1 & \text{if } e = t_2, \\ 0, & \text{otherwise.} \end{cases}$$

Now, using Proposition 3.9, it is sufficient to check if $\operatorname{Ker}\Phi \cap \operatorname{Ker} D^*_J$ intersect trivially for $J = \hat{J} = \operatorname{supp}(D^*\hat{x})$ and $J = \tilde{J} = \operatorname{supp}(D^*\tilde{x})$, which can be done by hand or using again an SVD of $\begin{pmatrix} \Phi \\ D^*_J \end{pmatrix}$. This leads to $X = \operatorname{conv}\{\hat{x}, \tilde{x}\} \ni \bar{x}$.

We illustrate in Figure 3.2 this construction when $n = 256$ and where the solution are obtained by the Chambolle–Pock algorithm [9]; the SVD and linear programming are computed with `scipy` [37].

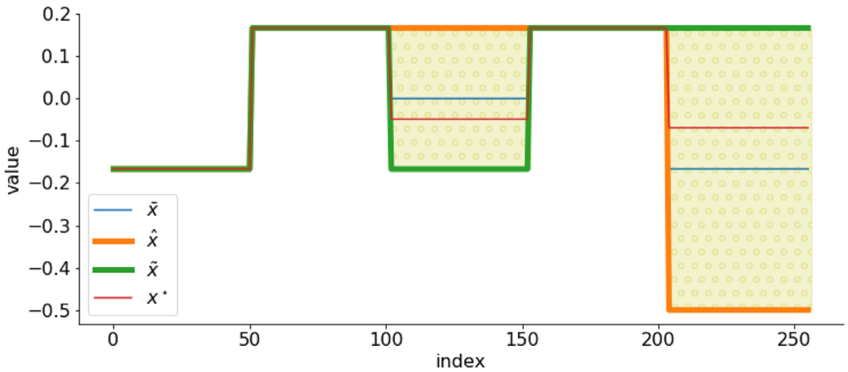

FIG. 3.2. *The solution set X represented as the convex hull of $\tilde{x}$ and $\hat{x}$.* (*Blue*) *Proposed solution $\bar{x}$.* (*Orange*) *Candidate extreme point $\hat{x}$.* (*Green*) *Computed extreme solution.* (*Red*) *Solution with Chambolle–Pock initialized with a random vector.* (*Yellow*) *Representation of the convex hull.*

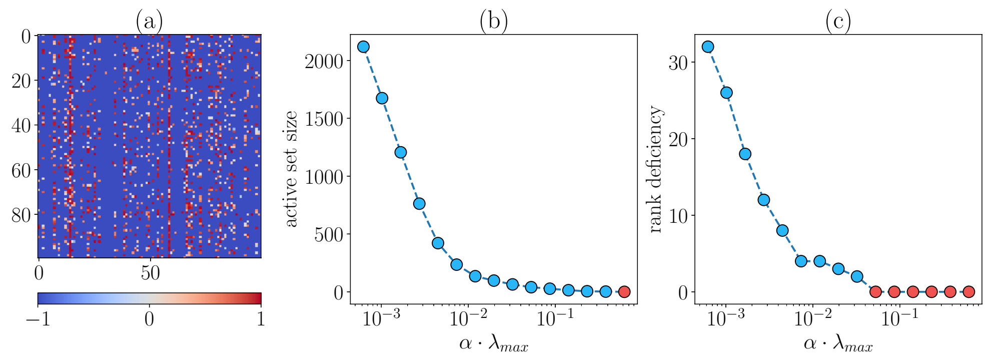

FIG. 3.3. *Lasso regularization on the* **gisette** *data set for various values of λ.* (a) *First* 100 *lines* (*total* 6000) *and columns* (*total* 5000) *of the matrix Φ in the data set. As observed on this sample, around* 12% *of the entries are positive.* (b) *Size of the support I with respect to the regularization parameter.* (c) *Difference between the size of the support I and the rank of Φ_I. Red points in* (a) *and* (b) *correspond to the value* 0.

3.6. Illustration on a real data set. We now consider the Lasso case where $D = \mathrm{Id}$. We consider the data set `gisette` that is available as a `libsvm` data set.[1] The `gisette` data set was introduced in the NeurIPS 2003 feature selection challenge [13]. This data set has $q = 5000$ samples and $n = 6000$ features. Its main characteristic is that most entries are ± 1 leading to a rank deficiency. Figure 3.3 shows the rank deficiency of a Lasso problem solved for various proportions α of $\lambda_{\max} = \|\Phi^* y\|_\infty$. This figure is generated with a FISTA solver [3] for $2 \cdot 10^4$ iterations in order to have a high accuracy solution (and starting from 0).[2] Such a behavior is not observed for generic data sets (e.g., it is not occurring for `20news,` for instance).

Solving the Lasso problem for $\lambda = \frac{1}{50}\lambda_{\max}$ leads to a solution $x^\star$ having an active set of size 134 but such that Φ_I has rank 130 (where $I = \mathrm{supp}(x^\star)$). It is then possible to construct an extreme point using the procedure described in the Appendix A.3 in

[1]Or at the following url: `https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/binary/gisette_scale.bz2`.

[2]The source code of this experiment is available as a gist: `https://gist.github.com/svaiter/e44ee3042a116580aaf33ca48bb4535b`.

[35]. For the sake of clarity, we recall this "H_J-procedure": if the support I of $x^\star$ is such that Φ_I does not have full rank,

1. Take $h \in \operatorname{Ker} \Phi_I$;
2. Consider the vectors $x_t = x^\star + th$ for $t > 0$. There exists a supremum t_0 such that x_t is a solution, and one can prove that $I_0 = \operatorname{supp}(x_{t_0}) \subset I$.

Iterating this procedure leads to solution with full-rank since the size of the support decreases at least by one at each iteration.

4. Conclusion. In this work, we have refined the analysis of the solution set of sparse ℓ^1 analysis regularization to understand its geometry. To perform this analysis, we have drawn an explicit relationship between the structure of the unit ball of the regularizer and the set of feasible signs. Upon this work, we derived a necessary and sufficient condition for a convex set to be the solution of sparse analysis regularization problem. Extension of our results to nonconvex sparse analysis penalizations such as $\|\cdot\|_p$ with $0 < p < 1$ is an interesting research direction, where face decomposition of the polytope unit-ball needs to be replaced with stratification of semialgebraic sets.

From a practical point of view, this work adds another argument towards the need for a good choice of regularizer/dictionary when a user seeks a robust and unique solution to its optimization problem. This work is mainly of theoretical interest since numerical applications should deal with exponential algorithms with respect to the signal dimension. Note, however, that in the case of the expected sparsity level when the maximal solution is logarithmic in the dimension, the enumeration problem is in this case tractable. We believe that the results contained in this paper will help other theoretical works around sparse analysis regularization, such as performing the sensitivity analysis of (1.1) with respect to the dictionary used in the regularization.

Acknowledgments. The authors thank P. Tardivel for pointing out dual geometrical conditions for uniqueness and M. Massias for suggesting to use the `gisette` data set to illustrate our results. We also thank the anonymous referees for their valuable comments.

REFERENCES

[1] A. Ali and R. J. Tibshirani, *The Generalized Lasso Problem and Uniqueness*, preprint, arXiv:1805.07682, 2018.

[2] A. Barbara, A. Jourani, and S. Vaiter, *Maximal solutions of sparse analysis regularization*, J. Optimiz. Theory Appl., 180 (2019), pp. 374–396, https://doi.org/10.1007/s10957-018-1385-3.

[3] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[4] L. J. Billera and C. W. Lee, *Sufficiency of McMullen's conditions for f-vectors of simplicial polytopes*, Bull. Amer. Math. Soc. (N.S.), 2 (1980), pp. 181–185.

[5] A. Björner, M. Las Vergnas, B. Sturmfels, N. White, and G. M. Ziegler, *Oriented Matroids*, Vol. 46, Cambridge University Press, 1999.

[6] E. D. Bolker, *A class of convex bodies*, Trans. Am. Math. Soc., 145 (1969), pp. 323–345.

[7] C. Boyer, A. Chambolle, Y. Castro, V. Duval, F. de Gournay, and P. Weiss, *On representer theorems and convex regularization*, SIAM J. Optimiz., 29 (2019), pp. 1260–1281, https://doi.org/10.1137/18M1200750.

[8] J. Burke and M. C. Ferris, *Characterization of solution sets of convex programs*, Oper. Res. Lett., 10 (1991), pp. 57–60.

[9] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Vis. Imaging, 40 (2011), pp. 120–145.

[10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, *Least angle regression*, Ann. Stat., 32 (2004), pp. 407–499, https://doi.org/10.1214/009053604000000067.

[11] M. Elad, P. Milanfar, and R. Rubinstein, *Analysis versus synthesis in signal priors*, Inverse Probl., 23 (2007), pp. 947–968, https://doi.org/10.1088/0266-5611/23/3/007.

[12] K. Ewald and U. Schneider, *On the distribution, model selection properties and uniqueness of the lasso estimator in low and high dimensions*, Electron. J. Stat., 14 (2020), pp. 944–969.

[13] I. Guyon, S. Gunn, A. Ben-Hur, and G. Dror, *Result analysis of the NIPS* 2003 *feature selection challenge*, NeurIPS, 17 (2004).

[14] J.-B. Hiriart-Urruty and C. Lemaréchal, *Convex analysis and minimization algorithms* I*: Fundamentals*, Vol. 305, Springer Science & Business Media, New York, 2013.

[15] V. Jeyakumar and X. Yang, *On characterizing the solution sets of pseudolinear programs*, J. Optimiz. Theory Appl., 87 (1995), pp. 747–755.

[16] G. Kimeldorf and G. Wahba, *Some results on Tchebycheffian spline functions*, J. Math. Anal. Appl., 33 (1971), pp. 82–95, https://doi.org/10.1016/0022-247X(71)90184-3.

[17] J. Mairal and B. Yu, *Complexity Analysis of the Lasso Regularization Path*, preprint, arXiv:1205.0079, 2012.

[18] O. Mangasarian, *Minimum-support solutions of polyhedral concave programs*, Optimization, 45 (1999), pp. 149–162.

[19] O. L. Mangasarian, *A simple characterization of solution sets of convex programs*, Oper. Res. Lett., 7 (1988), pp. 21–26.

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

[21] S. Nam, M. E. Davies, M. Elad, and R. Gribonval, *The cosparse analysis model and algorithms*, Appl. Comput. Harmon. Anal., 34 (2013), pp. 30–56, https://doi.org/10.1016/j.acha.2012.03.006.

[22] R. T. Rockafellar, *Convex Analysis*, Princeton Math. Ser. 28, Princeton University Press, Princeton, NJ, 1970.

[23] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D Nonlinear Phenom., 60 (1992), pp. 259–268. https://doi.org/10.1016/0167-2789(92)90242-F.

[24] U. Schneider and P. Tardivel, *The Geometry of Uniqueness, Sparsity and Clustering in Penalized Estimation*, preprint, arXiv:2004.09106, 2020.

[25] B. Schölkopf, R. Herbrich, and A. J. Smola, *A generalized representer theorem*, in International Conference on Computational Learning Theory, Springer, New York, 2001, pp. 416–426.

[26] J. Sharpnack, A. Singh, and A. Rinaldo, *Sparsistency of the edge lasso over graphs*, in Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics, N. D. Lawrence and M. Girolami, eds., Proc. Mach. Learn. Res. 22, La Palma, Canary Islands, PMLR, 2012, pp. 1028–1036, https://proceedings.mlr.press/v22/sharpnack12.html.

[27] Y. She, *Sparse regression with exact clustering*, Electron. J. Stat., 4 (2010), pp. 1055–1096, https://doi.org/10.1214/10-EJS578.

[28] G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk, *On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides*, SIAM J. Numer. Anal., 42 (2004), pp. 686–713, https://doi.org/10.1137/S0036142903422429.

[29] P. J. C. Tardivel, T. Skalski, P. Graczyk, and U. Schneider, *The Geometry of Model Recovery by Penalized and Thresholded Estimators*, in process, 2021, https://hal.archives-ouvertes.fr/hal-03262087.

[30] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. R. Stat. Soc. Ser. B (Methodol.), 58 (1996), pp. 267–288.

[31] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, *Sparsity and smoothness via the fused lasso*, J. R. Stat. Soc. Ser. B (Stat. Methodol.), 67 (2005), pp. 91–108, https://doi.org/10.1111/j.1467-9868.2005.00490.x.

[32] R. J. Tibshirani, *The lasso problem and uniqueness*, Electron. J. Statist., 7 (2013), pp. 1456–1490, https://doi.org/10.1214/13-EJS815.

[33] R. J. Tibshirani and J. Taylor, *The solution path of the generalized lasso*, Ann. Statist., 39 (2011), pp. 1335–1371, https://doi.org/10.1214/11-AOS878.

[34] M. Unser, *A Unifying Representer Theorem for Inverse Problems and Machine Learning*, preprint, arXiv:1903.00687, 2019.

[35] S. Vaiter, C. Deledalle, G. Peyré, C. Dossal, and J. M. Fadili, *Local behavior of sparse analysis regularization: Applications to risk estimation*, Appl. Comput. Harmon. Anal., 35 (2013), pp. 433–451, https://doi.org/10.1016/j.acha.2012.11.006.

[36] S. Vaiter, G. Peyre, C. Dossal, and J. Fadili, *Robust sparse analysis regularization*, IEEE Trans. Inform. Theory, 59 (2013), pp. 2001–2016, https://doi.org/10.1109/TIT.2012.2233859.

[37] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, et al, *SciPy* 1.0: *Fundamental algorithms for scientific computing in python*, Nat. Methods, 17 (2020), pp. 261–272, https://doi.org/10.1038/s41592-019-0686-2.

[38] H. Zhang, M. Yan, and W. Yin, *One condition for solution uniqueness and robustness of both l1-synthesis and* l1-*analysis minimizations*, Adv. Comput. Math., 42 (2016), pp. 1381–1399, https://doi.org/10.1007/s10444-016-9467-y.

[39] G. M. Ziegler, *Lectures on Polytopes*, Vol. 152, Springer Science & Business Media, New York, 1995.