Real vs. Complex Null Space Properties for Sparse Vector Recovery *

Simon Foucart, Rémi Gribonval

Abstract

We identify and solve an overlooked problem about the characterization of underdetermined systems of linear equations for which sparse solutions have minimal ℓ_1-norm. This characterization is known as the *null space property*. When the system has real coefficients, sparse solutions can be considered either as real or complex vectors, leading to two seemingly distinct null space properties. We prove that the two properties actually coincide by establishing a link with a problem about convex polygons in the real plane. Incidentally, we also show the equivalence between *stable null space properties* which account for the stable reconstruction by ℓ_1-minimization of vectors that are not exactly sparse.

Nous identifions et résolvons un problème lié aux systèmes sous-determinés d'équations linéaires, plus précisément à la propriété de leurs noyaux qui caractérise le fait que les solutions parcimonieuses soient celles avec la plus petite norme ℓ_1. Quand les coefficients du système sont réels, les solutions parcimonieuses peuvent être considérés comme vecteurs réels ou complexes, ce qui conduit à deux propriétés des noyaux a priori distinctes. Nous démontrons que ces deux propriétés sont en fait équivalentes en établissant un lien avec un problème sur les polygones convexes du plan réel. Accessoirement, nous prouvons aussi l'équivalence entre des propriétés *stables* du noyau, lesquelles expliquent la stabilité de la reconstruction par minimisation ℓ_1 de vecteurs qui ne sont pas exactement parcimonieux.

This note deals with the recovery of sparse vectors $\mathbf{x}$ from incomplete measurements $\mathbf{y} = A\mathbf{x}$, where A is an $m \times N$ matrix with $m \ll N$. The interest in developing sparse data models for solving ill-posed inverse problems originates in the possibility of such a recovery in underdetermined situations. This is also the fundamental result underlying the recent field of Compressed Sensing, which aims at acquiring signals/images well below the Nyquist rate by exploiting their sparsity in an appropriate domain. It is well known by now that the recovery can be carried out by solving the convex optimization problem

$$(\mathrm{P}_1) \qquad \text{minimize } \ \|\mathbf{z}\|_1 \quad \text{subject to } A\mathbf{z} = \mathbf{y},$$

*This work has been supported by the French National Research Agency (ANR) through the project ECHANGE (ANR-08-EMER-006).

provided some suitable conditions are satisfied by the matrix A — actually, by its null space. To be more precise, see [2, 3] for details, every vector $\mathbf{x}$ supported on a set S is the unique solution of (P_1) with $\mathbf{y} = A\mathbf{x}$ if and only if

$$\|\mathbf{u}_S\|_1 < \|\mathbf{u}_{\overline{S}}\|_1 \quad \text{for all } \mathbf{u} \in \ker A \setminus \{0\}. \tag{1}$$

The set $\overline{S}$ designates the complementary of the set S, and the notations $\mathbf{u}_S$ and $\mathbf{u}_{\overline{S}}$ stand for the vectors whose entries indexed by S and $\overline{S}$, respectively, equal those of $\mathbf{u}$, while the other entries are set to zero. We are also interested in a strengthening of Property (1), namely

$$\|\mathbf{u}_S\|_1 \leq \rho \, \|\mathbf{u}_{\overline{S}}\|_1 \quad \text{for all } \mathbf{u} \in \ker A \setminus \{0\}, \tag{2}$$

for some $0 < \rho < 1$. This *stable null space property* is actually equivalent to the property that

$$\|\mathbf{z} - \mathbf{x}\|_1 \leq \frac{1+\rho}{1-\rho} \Big[\|\mathbf{z}\|_1 - \|\mathbf{x}\|_1 + 2\|\mathbf{x}_{\overline{S}}\|_1 \Big] \quad \text{whenever } A\mathbf{z} = A\mathbf{x}.$$

Note that the latter implies that $\mathbf{z} = \mathbf{x}$ if $\mathbf{x}$ is supported on S and if $\mathbf{z}$ is a solution of (P_1) with $\mathbf{y} = A\mathbf{x}$. So far, we have deliberately been ambiguous about the underlying scalar field — often, this is not alluded at all. In fact, the above-mentioned equivalences are valid in the real and complex settings alike. However, since a real-valued measurement matrix A is also a complex-valued one, Properties (1) and (2) can be interpreted in two different ways. The real versions read

$$\sum_{j \in S} |u_j| < \sum_{\ell \in \overline{S}} |u_\ell| \quad \Big[\text{or} \leq \rho \sum_{\ell \in \overline{S}} |u_\ell| \Big] \quad \text{for all } \mathbf{u} \in \ker_{\mathbb{R}} A \setminus \{0\}, \tag{3}$$

while the complex versions read, in view of $\ker_{\mathbb{C}} A = \ker_{\mathbb{R}} A + i \ker_{\mathbb{R}} A$,

$$\sum_{j \in S} \sqrt{v_j^2 + w_j^2} < \sum_{\ell \in \overline{S}} \sqrt{v_\ell^2 + w_\ell^2} \quad \Big[\text{or} \leq \rho \sum_{\ell \in \overline{S}} \sqrt{v_\ell^2 + w_\ell^2} \Big] \quad \text{for all } (\mathbf{v}, \mathbf{w}) \in (\ker_{\mathbb{R}} A)^2 \setminus \{(0,0)\}. \tag{4}$$

We are going to show that Properties (3) and (4) are identical. Thus, every complex vector supported on S is recovered by ℓ_1-minimization if and only if every real vector supported on S is recovered by ℓ_1-minimization — informally, real and complex ℓ_1-recoveries succeed simultaneously. Before stating the theorem, we point out that, for a real measurement matrix, one may also recover separately the real and imaginary parts of a complex vector using two real ℓ_1-minimizations — which are linear programs — rather than recovering the vector directly using one complex ℓ_1-minimization — which is a second order cone program.

Theorem 1. For a measurement matrix $A \in \mathbb{R}^{m \times N}$ and a set $S \subseteq \{1, \ldots, N\}$, the real null space property (3) is equivalent to the complex null space property (4).

Proof. It is clear that (4) implies (3). Now, in order to handle null space properties and stable null space properties at once, we introduce the shorthand '$\prec$' to mean either '$<$' or '$\leq \rho$'. We assume that (3) holds, and we consider $(\mathbf{v}, \mathbf{w}) \in (\ker_{\mathbb{R}} A)^2 \setminus \{(0,0)\}$. We suppose that $\mathbf{v}$ and $\mathbf{w}$ are linearly independent, for otherwise (4) is clear. By applying (3) to $\mathbf{u} = \alpha \mathbf{v} + \beta \mathbf{w}$, we have

$$\sum_{j \in S} |\alpha v_j + \beta w_j| \prec \sum_{\ell \in \overline{S}} |\alpha v_\ell + \beta w_\ell| \quad \text{for all } \boldsymbol{\lambda} := (\alpha, \beta) \in \mathbb{R}^2 \setminus \{0\}.$$

If B_S and $B_{\overline{S}}$ denote the $2 \times s$ and $2 \times (N-s)$ matrices with columns $\mathbf{b}_j := [v_j, w_j]^\top$, $j \in S$, and $\mathbf{b}_\ell := [v_\ell, w_\ell]^\top$, $\ell \in \overline{S}$, respectively, this translates into

$$\|B_S^\top \boldsymbol{\lambda}\|_1 \prec \|B_{\overline{S}}^\top \boldsymbol{\lambda}\|_1 \quad \text{for all } \boldsymbol{\lambda} \in \mathbb{R}^2 \setminus \{0\}, \tag{5}$$

in other words

$$|\langle \boldsymbol{\lambda}, B_S \boldsymbol{\mu} \rangle| = |\langle B_S^\top \boldsymbol{\lambda}, \boldsymbol{\mu} \rangle| \prec \|B_{\overline{S}}^\top \boldsymbol{\lambda}\|_1 \quad \text{for all } \boldsymbol{\lambda} \in \mathbb{R}^2 \setminus \{0\} \text{ and all } \boldsymbol{\mu} \in \mathbb{R}^s \text{ with } \|\boldsymbol{\mu}\|_\infty = 1. \tag{6}$$

Let us observe that (5) implies the injectivity of $B_{\overline{S}}^\top$, hence the surjectivity of $B_{\overline{S}}$. Thus, for $\boldsymbol{\mu} \in \mathbb{R}^s$ with $\|\boldsymbol{\mu}\|_\infty = 1$, there exists $\boldsymbol{\nu} \in \mathbb{R}^{N-s}$ such that $B_{\overline{S}} \boldsymbol{\nu} = B_S \boldsymbol{\mu}$. As a result of (6), we have $|\langle B_{\overline{S}}^\top \boldsymbol{\lambda}, \boldsymbol{\nu} \rangle| \prec \|B_{\overline{S}}^\top \boldsymbol{\lambda}\|_1$ for all $\boldsymbol{\lambda} \in \mathbb{R}^2 \setminus \{0\}$. This means that the linear functional f defined on $\operatorname{ran} B_{\overline{S}}^\top$ by $f(\boldsymbol{\eta}) := \langle \boldsymbol{\eta}, \boldsymbol{\nu} \rangle$ has norm $\|f\|_1^* \prec 1$. By the Hahn–Banach theorem, we extend it to a linear functional $\widetilde{f}$ defined on the whole of $\mathbb{R}^{N-s}$. The latter can be represented as $\widetilde{f}(\boldsymbol{\eta}) = \langle \boldsymbol{\eta}, \widetilde{\boldsymbol{\nu}} \rangle$. The equality $\|\widetilde{f}\|_1^* = \|f\|_1^*$ translates into $\|\widetilde{\boldsymbol{\nu}}\|_\infty \prec 1$, while the identity $\widetilde{f}(\boldsymbol{\eta}) = f(\boldsymbol{\eta})$ for all $\boldsymbol{\eta} \in \operatorname{ran} B_{\overline{S}}^\top$ yields $0 = \langle B_{\overline{S}}^\top \boldsymbol{\lambda}, \widetilde{\boldsymbol{\nu}} - \boldsymbol{\nu} \rangle = \langle \boldsymbol{\lambda}, B_{\overline{S}} \widetilde{\boldsymbol{\nu}} - B_{\overline{S}} \boldsymbol{\nu} \rangle$ for all $\boldsymbol{\lambda} \in \mathbb{R}^2$, so that $B_{\overline{S}} \widetilde{\boldsymbol{\nu}} = B_{\overline{S}} \boldsymbol{\nu} = B_S \boldsymbol{\mu}$. In short, for any $\boldsymbol{\mu} \in \mathbb{R}^s$ with $\|\boldsymbol{\mu}\|_\infty = 1$, there exists $\widetilde{\boldsymbol{\nu}} \in \mathbb{R}^{N-s}$ with $\|\widetilde{\boldsymbol{\nu}}\|_\infty \prec 1$ such that $B_{\overline{S}} \widetilde{\boldsymbol{\nu}} = B_S \boldsymbol{\mu}$. Therefore, the convex polygon $\mathcal{C}_S := B_S[-1,1]^s$ is strictly contained in the convex polygon $\mathcal{C}_{\overline{S}} := B_{\overline{S}}[-1,1]^{N-s}$, respectively is contained in $\rho\, \mathcal{C}_{\overline{S}}$. This intuitively implies that

$$\operatorname{Perimeter}(\mathcal{C}_S) \prec \operatorname{Perimeter}(\mathcal{C}_{\overline{S}}). \tag{7}$$

In fact, the perimeter of a convex polygon $\mathcal{C}$ is the unique minimum of all the perimeters of compact convex sets containing the vertices of $\mathcal{C}$. This can be seen by isolating the contribution to the perimeter of each angular sector originating from a point inside $\mathcal{C}$ and intercepting two consecutive vertices. One can also invoke Cauchy's formula, see e.g. [6], for the perimeter of a compact convex set $\mathcal{K}$ as the integral of the length of the projection of $\mathcal{K}$ onto a line of direction θ, namely

$$\operatorname{Perimeter}(\mathcal{K}) = \int_0^\pi \Big[\max_{(x,y) \in \mathcal{K}} \big(x \cos(\theta) + y \sin(\theta) \big) - \min_{(x,y) \in \mathcal{K}} \big(x \cos(\theta) + y \sin(\theta) \big) \Big] d\theta.$$

The convex polygons $\mathcal{C}_S$ and of $\mathcal{C}_{\overline{S}}$ both take the form $M[-1,1]^n$, which may be viewed as the Minkowski sum of the line segments $[-\mathbf{c}_1, \mathbf{c}_1], \ldots, [-\mathbf{c}_n, \mathbf{c}_n]$, where $\mathbf{c}_1, \ldots, \mathbf{c}_n$ are the columns of the matrix M — in general dimension, Minkowski sums of line segments are called zonotopes. The perimeter of such convex polygons is explicitly given by $4(\|\mathbf{c}_1\|_2 + \cdots + \|\mathbf{c}_n\|_2)$, see e.g. [5]. Thus, (7) reads

$$4 \sum_{j \in S} \|\mathbf{b}_j\|_2 \prec 4 \sum_{\ell \in S} \|\mathbf{b}_\ell\|_2.$$

Up to the factor 4, this is the complex null space property (4). □

Remark. Sparse recovery can also be achieved by ℓ_q-minimization for $0 < q < 1$. Its success on a set S is characterized by the null space property $\|\mathbf{u}_S\|_q < \|\mathbf{u}_{\overline{S}}\|_q$ for all $\mathbf{u} \in \ker A \setminus \{0\}$, see [3]. The same ambiguity about its real or complex interpretation arises. The question whether the two notions coincide in this case is open.

Remark. The recovery of sparse complex vectors by ℓ_1-minimization can be viewed as the special case $n = 2$ of the recovery of jointly sparse real vectors by mixed $\ell_{1,2}$-minimization. In this context, see [1] for details, every n-tuple $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ of vectors in $\mathbb{R}^N$, each of which supported on the same set S, is the unique solution of

$$\text{minimize} \quad \sum_{j=1}^{N} \sqrt{z_{1,j}^2 + \cdots + z_{n,j}^2} \quad \text{subject to } A\mathbf{z}_1 = A\mathbf{x}_1, \ldots, A\mathbf{z}_n = A\mathbf{x}_n,$$

if and only if a mixed $\ell_{1,2}$ null space property holds, namely

$$(8) \quad \sum_{j \in S} \sqrt{u_{1,j}^2 + \cdots + u_{n,j}^2} < \sum_{\ell \in \overline{S}} \sqrt{u_{1,\ell}^2 + \cdots + u_{n,\ell}^2} \quad \text{for all } (\mathbf{u}_1, \ldots, \mathbf{u}_n) \in \big(\ker_{\mathbb{R}} A\big)^n \setminus \{(0, \ldots, 0)\}.$$

It is then natural to wonder whether the real null space property (3) implies the mixed $\ell_{1,2}$ null space property (8) when $n \geq 3$. This is also an open question.

Added in proof. Since the submission of this note, the two open questions raised in the remarks have been answered in the affirmative by Lai and Liu [4].

References

[1] E. van den Berg, M. P. Friedlander, Joint-sparse recovery from multiple measurements, preprint.

[2] D. Donoho, M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization. Proc. Natl. Acad. Sci. USA 100(5) (2003), 2197–2202.

[3] R. Gribonval, M. Nielsen, Highly sparse representations from dictionaries are unique and independent of the sparseness measure, Appl. Comput. Harm. Anal., 22(3) (2007), 335–355.

[4] M.-J. Lai, Y. Liu, The null space property for sparse recovery from multiple measurement vectors, preprint.

[5] S. Markov, On quasivector spaces of convex bodies and zonotopes, Numer. Algorithms, 37 (2004), 263–274.

[6] F. Spitzer, H. Widom, The circumference of a convex polygon, Proceedings of the American Mathematical Society, 12(3) (1961), 506–509.