

# Supplementary material

Patrick J.C. Tardivel<sup>a,\*</sup> and Małgorzata Bogdan<sup>a,b</sup>,

<sup>a</sup> Institute of Mathematics, University of Wrocław, Wrocław, Poland

<sup>b</sup> Department of Statistics, Lund University, Lund, Sweden

## General position condition

Let  $M_1, \dots, M_k \in \mathbb{R}^p$  and let  $\text{Aff}\{M_i\}_{1 \leq i \leq k}$  denotes for the following affine subspace

$$\text{Aff}\{M_i\}_{1 \leq i \leq k} := \{M \in \mathbb{R}^p \mid M = \alpha_1 M_1 + \dots + \alpha_k M_k \text{ where } \alpha_1, \dots, \alpha_k \in \mathbb{R} \text{ and } \alpha_1 + \dots + \alpha_k = 1\}.$$

The definition of general position given in Tibshirani [4] is as follows.

**Definition 1 (General position)** *The  $n \times p$  matrix  $A = (A_1 | \dots | A_p)$  is in general position if whatever  $s \in \{-1, 1\}^p$ , whatever  $I \subset \{1, \dots, p\}$  such that  $\text{card}(I) \leq \min\{n, p\}$ , the affine space  $\text{Aff}\{s_i A_i\}_{i \in I}$  satisfies the following property*

$$\forall j \notin I, s_j A_j \notin \text{Aff}\{s_i A_i\}_{i \in I}.$$

**Example:** Let  $A = (A_1 | A_2 | A_3)$  and  $B = (B_1 | B_2 | B_3)$  be the following  $2 \times 3$  matrices

$$A = \begin{pmatrix} 1 & 0 & -1/2 \\ 0 & 1 & -1/2 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

Matrix  $A$  is not in general position since points  $A_1, A_2$  and  $-A_3$  are aligned. Matrix  $B$  is in general position because, up to a sign,  $B_1, B_2$  and  $B_3$  are not aligned.

Proposition 1 shows that when  $X$  is in general position then the BPDN problem has a unique minimizer.

**Proposition 1** *Let  $X$  be a  $n \times p$  matrix in general position, let  $y$  be an arbitrary element of  $\mathbb{R}^n$  and let  $R \geq 0$ .*

*Let  $S_1^R$  be the solutions of the following optimization problem*

$$S_1^R := \underset{\gamma \in \mathbb{R}^p}{\text{argmin}} \|\gamma\|_1 \text{ subject to } \|y - X\gamma\|_2^2 \leq R.$$

---

\*Corresponding author: tardivel@math.uni.wroc.pl

If  $S_1^R$  is not empty then  $S_1^R$  is a singleton.

When  $\mathbf{0}$  is an element of  $S_1^R$ , then  $S_1^R = \{\mathbf{0}\}$  and thus proposition 1 holds. For this reason, hereafter, it is assumed that  $\mathbf{0} \notin S_1^R$ . Proposition 1 is a straightforward implication of Lemmas 1 and 2.

**Lemma 1** *Let  $X = (X_1 | \dots | X_p)$  be a  $n \times p$  matrix in general position and let  $S_1^R$  as defined in Proposition 1. If  $S_1^R$  is not empty then whatever  $\bar{\gamma} \in S_1^R$ , the family  $(X_i)_{i \in \text{supp}(\bar{\gamma})}$  is linearly independent.*

**Proof:** To simplify the notation, without any loss of generality let us assume that  $\text{supp}(\bar{\gamma}) = \{1, \dots, k_0\}$ . Let us assume that the family  $(X_i)_{1 \leq i \leq k_0}$  is not linearly independent. What follows is a contradiction for  $X$  in general position.

Let  $b = X\bar{\gamma}$  and let  $\bar{s} = (\text{sign}(\bar{\gamma}_i))_{1 \leq i \leq k_0}$ . Because  $\bar{\gamma} \in S_1^R$ , the vector  $(\bar{\gamma}_1, \dots, \bar{\gamma}_{k_0})$  is a solution of the following problem

$$\underset{\gamma \in \mathbb{R}^{k_0}}{\text{argmin}} \underbrace{\sum_{i=1}^{k_0} |\gamma_i|}_{\phi_0(\gamma_1, \dots, \gamma_{k_0})} \quad \text{subject to the constraint } \forall i \in \{1, \dots, n\}, \underbrace{\sum_{j=1}^{k_0} \gamma_j X_{ij}}_{\phi_i(\gamma_1, \dots, \gamma_{k_0})} = b_i.$$

Because  $\phi_0$  is differentiable at  $(\bar{\gamma}_1, \dots, \bar{\gamma}_{k_0})$  (since  $\bar{\gamma}_1 \neq 0, \dots, \bar{\gamma}_{k_0} \neq 0$ ), by the Lagrange multipliers theorem the following implications hold (where  $\nabla$  denotes for the gradient).

$$\begin{aligned} & \nabla \phi_0(\bar{\gamma}_1, \dots, \bar{\gamma}_{k_0}) \in \text{span} \{ \nabla \phi_1(\bar{\gamma}_1, \dots, \bar{\gamma}_{k_0}), \dots, \nabla \phi_n(\bar{\gamma}_1, \dots, \bar{\gamma}_{k_0}) \}, \\ \Rightarrow & \exists \lambda = (\lambda_1, \dots, \lambda_n) \text{ such that } (\bar{s}_1, \dots, \bar{s}_{k_0}) = \lambda_1 (X_{11}, \dots, X_{1k_0}) + \dots + \lambda_n (X_{n1}, \dots, X_{nk_0}), \\ \Rightarrow & \forall i \in \{1, \dots, k_0\}, \bar{s}_i = \lambda_1 X_{1i} + \dots + \lambda_n X_{ni} = X_i' \lambda \end{aligned} \quad (1)$$

Let  $(X_i)_{i \in I}$  be a basis of the vectorial space  $\text{span} \{X_1, \dots, X_{k_0}\}$  (since  $(X_i)_{i \in I}$  is a basis of a subspace of  $\mathbb{R}^n$ , then  $\text{card}(I) \leq n$ , furthermore  $\text{card}(I) \leq k_0 \leq p$ ). Because the family  $(X_i)_{1 \leq i \leq k_0}$  is not then linearly independent, there exists  $j \in \{1, \dots, k_0\} \setminus I$  and  $(\alpha_i)_{i \in I}$  such that  $X_j = \sum_{i \in I} \alpha_i X_i$ . The following implication is a straightforward consequence of (1)

$$\bar{s}_j = X_j' \lambda = \sum_{i \in I} \alpha_i X_i' \lambda = \sum_{i \in I} \alpha_i \bar{s}_i \Rightarrow 1 = (\bar{s}_j)^2 = \sum_{i \in I} \alpha_i \bar{s}_i \bar{s}_j.$$

Finally, the following equality

$$X_j = \sum_{i \in I} \alpha_i X_i = \sum_{i \in I} \alpha_i (\bar{s}_j \bar{s}_i)^2 X_i = \sum_{i \in I} \alpha_i \bar{s}_j \bar{s}_i (\bar{s}_i \bar{s}_j X_i) \quad (2)$$

shows that  $X_j$  is the barycentre of the family  $(\bar{s}_i \bar{s}_j X_i)_{i \in I}$  since  $1 = \sum_{i \in I} \alpha_i \bar{s}_i \bar{s}_j$ . Consequently, equality (2) contradicts that  $X$  is in general position.  $\square$

**Lemma 2** Let  $S_1^R$  as defined in Proposition 1 and let us assume that whatever  $\bar{\gamma} \in S_1^R$  the family  $(X_i)_{i \in \text{supp}(\bar{\gamma})}$  is linearly independent then  $S_1^R$  is a singleton.

**Proof:** Let  $\bar{\gamma}^1$  and  $\bar{\gamma}^2$  be two elements of  $S_1^R$ . In the first step, let us show that  $X\bar{\gamma}^1 = X\bar{\gamma}^2$ . Since function  $t \in \mathbb{R}^p \mapsto \|y - t\|_2^2$  is strictly convex, if  $X\bar{\gamma}^1 \neq X\bar{\gamma}^2$  then whatever  $\alpha \in (0, 1)$  the following inequality holds

$$\|y - \alpha X\bar{\gamma}^1 - (1 - \alpha)X\bar{\gamma}^2\|_2^2 < \alpha\|y - X\bar{\gamma}^1\|_2^2 + (1 - \alpha)\|y - X\bar{\gamma}^2\|_2^2 \leq R.$$

Let us define  $\bar{\gamma}_\alpha := \alpha\bar{\gamma}^1 + (1 - \alpha)\bar{\gamma}^2$  where  $\alpha \in (0, 1)$ . Let us set, for example,  $\alpha = 1/2$  from the previous inequality. One may deduce that 1)  $\|y - X\bar{\gamma}_{1/2}\|_2^2 < R$ . Furthermore, since the  $l^1$  norm is convex, one may deduce that 2)  $\bar{\gamma}_{1/2} \in S_1^R$ . Let us show that 1) and 2) provide a contradiction. Since  $\|y - X\bar{\gamma}_{1/2}\|_2^2 < R$  then there exists  $t \in [0, 1)$  such that  $t\bar{\gamma}_{1/2}$  is still a feasible point, namely  $\|y - Xt\bar{\gamma}_{1/2}\|_2^2 \leq R$ . Because  $\bar{\gamma}_{1/2} \neq \mathbf{0}$  (since  $\mathbf{0} \notin S_1^R$ ), we have  $\|t\bar{\gamma}_{1/2}\|_1 < \|\bar{\gamma}_{1/2}\|_1$  which is contradictory with  $\bar{\gamma}_{1/2} \in S_1^R$ .

To conclude this proof it is sufficient to show that the family  $(X_i)_{i \in I}$  where  $I = \text{supp}(\bar{\gamma}_1) \cup \text{supp}(\bar{\gamma}_2)$  is linearly independent. It is straightforward that there exists  $\alpha \in (0, 1)$  such that  $I = \text{supp}(\bar{\gamma}_\alpha)$ . Because  $X\bar{\gamma}^1 = \sum_{i \in I} \bar{\gamma}_i^1 X_i = \sum_{i \in I} \bar{\gamma}_i^2 X_i = X\bar{\gamma}_2$  and because the family  $(X_i)_{i \in I}$  is linearly independent (since  $\bar{\gamma}_\alpha \in S_1^R$ ) then  $\bar{\gamma}_1 = \bar{\gamma}_2$ .  $\square$

Proposition 2 shows that once the BPDN (resp. BP) problem has a unique minimizer  $\bar{\gamma}$ , then the family  $(X_i)_{i \in \text{supp}(\bar{\gamma})}$  is linearly independent.

**Proposition 2** Let  $X$  be a  $n \times p$  matrix, let  $y$  be an arbitrary element of  $\mathbb{R}^n$  and let  $R \geq 0$ . Let us denote  $S_1^R$  be the solutions of the following optimization problem

$$S_1^R := \underset{\gamma \in \mathbb{R}^p}{\text{argmin}} \|\gamma\|_1 \text{ subject to } \|y - X\gamma\|_2^2 \leq R. \quad (3)$$

If  $S_1^R$  has a unique element  $\bar{\gamma} \neq \mathbf{0}$ , then the family  $(X_i)_{i \in \text{supp}(\bar{\gamma})}$  is linearly independent.

Let us assume that the family  $(X_i)_{i \in \text{supp}(\bar{\gamma})}$  is not linearly independent. We are going to provide a contradiction for the uniqueness of  $\bar{\gamma}$  by constructing a feasible point  $\tilde{\gamma}$  for which  $\tilde{\gamma} \neq \bar{\gamma}$  and  $\|\tilde{\gamma}\|_1 \leq \|\bar{\gamma}\|_1$ . Because  $(X_i)_{i \in \text{supp}(\bar{\gamma})}$  is not linearly independent, there exists coefficients  $(c_i)_{i \in \text{supp}(\bar{\gamma})}$  not simultaneously null such that  $\sum_{i \in \text{supp}(\bar{\gamma})} c_i X_i = \mathbf{0}$ . This equality provides an element  $h \in \ker(X) \setminus \mathbf{0}$  which is defined hereafter by

$$\forall i \in \text{supp}(\bar{\gamma}), h_i = c_i \text{ and } \forall i \notin \text{supp}(\bar{\gamma}), h_i = 0.$$

From this element  $h$ , one defines feasible points for problem (3) by setting for all  $t \in \mathbb{R}$ ,  $\gamma(t) = \bar{\gamma} + th$  (namely, whatever  $t \in \mathbb{R}$ ,  $\|y - X\gamma(t)\|_2^2 \leq R$ ). Let  $f$  be the function  $\forall t \in \mathbb{R}, f(t) := \|\gamma(t)\|_1$ .

Now, without loss of generality and to simplify the notation, let us assume that  $\text{supp}(h) = \{1, \dots, k_0\}$  thus,  $f(t) = \sum_{i=1}^{k_0} |\tilde{\gamma}_i + th_i| + \sum_{i=k_0+1}^p |\tilde{\gamma}_i|$ . Whatever  $i \in \{1, \dots, k_0\}$ , let us denote  $t_i = -\tilde{\gamma}_i/h_i$  (thus  $t_i \neq 0$ ), the function  $t \in \mathbb{R} \mapsto |\tilde{\gamma}_i + th_i|$  is coercive (*i.e.*  $\lim_{t \rightarrow +\infty} |\tilde{\gamma}_i + th_i| = \lim_{t \rightarrow -\infty} |\tilde{\gamma}_i + th_i| = +\infty$ ) and affine on the intervals  $(-\infty, t_i]$  and  $[t_i, \infty)$ . Consequently, the function  $f$  is coercive and affine on the intervals  $(-\infty, t_{(1)}], [t_{(1)}, t_{(2)}], \dots, [t_{(k_0)}, +\infty)$  where  $(\cdot)$  is a permutation such that  $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(k_0)}$ . Finally, the minimum of  $f$  is reached at  $t_0 \neq 0$  with  $t_0 \in \{t_1, \dots, t_{k_0}\}$ , thus there is a feasible point  $\tilde{\gamma} = \gamma(t_0)$  such that  $\tilde{\gamma} \neq \bar{\gamma}$  and  $\|\tilde{\gamma}\|_1 \leq \|\bar{\gamma}\|_1$  which provides a contradiction.  $\square$

## The stable nullspace property implies the identifiability condition

Let  $X$  be a  $n \times p$  matrix, given  $\rho \in (0, 1)$  the vector  $b \in \mathbb{R}^p$  satisfies the stable nullspace property if the following inequality occurs

$$\forall h \in \ker(X), \quad \sum_{i \in \text{supp}(b)} |h_i| \leq \rho \sum_{i \notin \text{supp}(b)} |h_i|.$$

One may notice that the stable nullspace property implies the following inequality

$$\forall h \in \ker(X) \setminus \{0\}, \quad \sum_{i \in \text{supp}(b)} \text{sign}(b_i) h_i < \sum_{i \notin \text{supp}(b)} |h_i|,$$

which characterizes that  $b$  is identifiable with respect to the  $L_1$  norm. Thus the stable nullspace property implies the identifiability condition. This statement is not new in the compressed sensing community. Actually, it is well known that the stable nullspace property implies  $L_1$  recovery by BP (see *e.g.* [1, 2, 3]).

Now, let us point out an exemple in which the identifiability condition holds but the stable nullspace condition does not hold. Let  $X$  and  $b$  be defined as follows

$$X := \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \quad \text{and} \quad b := \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

Because  $\ker(X)$  is spanned by the vector  $(1, 1, -1)$ , one may deduce that

$$\forall h \in \ker(X), \quad \sum_{i \in \text{supp}(b)} |h_i| = 2 \sum_{i \notin \text{supp}(b)} |h_i|,$$

thus the stable null space property does not hold. On the other hand,

$$\forall h \in \ker(X), \quad \sum_{i \in \text{supp}(b)} \text{sign}(b_i) h_i = 0,$$

thus the identifiability condition holds.

## References

- [1] David L Donoho and Michael Elad. Optimally sparse representation in general (nonorthogonal) dictionaries via  $\ell_1$  minimization. *Proceedings of the National Academy of Sciences*, 100(5):2197–2202, 2003.
- [2] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*, volume 1. Springer, 2013.
- [3] Rémi Gribonval and Morten Nielsen. Sparse representations in unions of bases. *IEEE Transactions on Information Theory*, 49(12):3320–3325, 2003.
- [4] Ryan J Tibshirani et al. The lasso problem and uniqueness. *Electronic Journal of Statistics*, 7:1456–1490, 2013.