Robust Polyhedral Regularization

Samuel Vaiter, Gabriel Peyré
CEREMADE, CNRS-Université Paris-Dauphine,
Place du Maréchal De Lattre De Tassigny,
75775 Paris Cedex 16, France.
Email: {vaiter,peyre}@ceremade.dauphine.fr

Jalal Fadili
GREYC, CNRS-ENSICAEN-Université de Caen,
6, Bd du Maréchal Juin,
14050 Caen Cedex, France.
Email: jalal.fadili@greyc.ensicaen.fr

***Abstract*—In this paper, we establish robustness to noise perturbations of polyhedral regularization of linear inverse problems. We provide a sufficient condition that ensures that the polyhedral face associated to the true vector is equal to that of the recovered one. This criterion also implies that the ℓ^2 recovery error is proportional to the noise level for a range of parameter. Our criterion is expressed in terms of the hyperplanes supporting the faces of the unit polyhedral ball of the regularization. This generalizes to an arbitrary polyhedral regularization results that are known to hold for sparse synthesis and analysis ℓ^1 regularization which are encompassed in this framework. As a byproduct, we obtain recovery guarantees for ℓ^∞ and $\ell^1 - \ell^\infty$ regularization.**

I. Introduction

A. Polyhedral Regularization

We consider the following linear inverse problem

$$y = \Phi x_0 + w, \tag{1}$$

where $y \in \mathbb{R}^Q$ are the observations, $x_0 \in \mathbb{R}^N$ is the unknown true vector to recover, w the bounded noise, and Φ a linear operator which maps the signal domain $\mathbb{R}^N$ into the observation domain $\mathbb{R}^Q$. The goal is to recover x_0 either exactly or to a good approximation.

We call a polyhedron a subset $\mathcal{P}$ of $\mathbb{R}^N$ such that $\mathcal{P} = \left\{x \in \mathbb{R}^N \mid Ax \leqslant b\right\}$ for some $A \in \mathbb{R}^{N_H \times N}$ and $b \in \mathbb{R}^{N_H}$, where the inequality $\leqslant$ should be understood component-wise. This is a classical description of convex polyhedral sets in terms of the hyperplanes supporting their $(N-1)$-dimensional faces.

In the following, we consider polyhedral convex functions of the form

$$J_H(x) = \max_{1 \leqslant i \leqslant N_H} \langle x,\, h_i \rangle,$$

where $H = (h_i)_{i=1}^{N_H} \in \mathbb{R}^{N \times N_H}$. Thus, $\mathcal{P}_H = \left\{x \in \mathbb{R}^N \mid J_H(x) \leqslant 1\right\}$ is a polyhedron. We assume that $\mathcal{P}_H$ is a bounded polyhedron which contains 0 in its interior. This amounts to saying that J_H is a gauge, or equivalently that it is continuous, non-negative, sublinear (i.e. convex and positively homogeneous), coercive, and $J_H(x) > 0$ for $x \neq 0$. Note that it is in general not a norm because it needs not be symmetric.

In order to solve the linear inverse problem (1), we devise the following regularized problem

$$x^\star \in \underset{x \in \mathbb{R}^N}{\operatorname{argmin}}\ \frac{1}{2}\|y - \Phi x\|^2 + \lambda J_H(x), \tag{$P_\lambda(y)$}$$

where $\lambda > 0$ is the regularization parameter. Coercivity and convexity of J_H implies the set of minimizers is non-empty, convex and compact.

In the noiseless case, $w = 0$, one usually considers the equality-constrained optimization problem

$$x^\star \in \underset{\Phi x = y}{\operatorname{argmin}}\ J_H(x). \tag{$P_0(y)$}$$

B. Relation to Sparsity and Anti-sparsity

Examples of polyhedral regularization include the ℓ^1-norm, analysis ℓ^1-norm and ℓ^∞-norm. The ℓ^1 norm reads

$$J_{H_1}(x) = \|x\|_1 = \sum_{i=1}^N |x_i|.$$

It corresponds to choosing $H_1 \in \mathbb{R}^{N \times 2^N}$ where the columns of H_1 enumerate all possible sign patterns of length N, i.e. $\{-1, 1\}^N$. The corresponding regularized problem $(P_\lambda(y))$ is the popular Lasso [1] or Basis Pursuit DeNoising [2]. It is used for recovering sparse vectors. Analysis-type sparsity-inducing penalties are obtained through the (semi-)norm $J_H(x) = \|Lx\|_1$, where $L \in \mathbb{R}^{P \times N}$ is an analysis operator. This corresponds to using $H = L^* H_1$ where * stands for the adjoint. A popular example is the anisotropic total variation where L is a first-order finite difference operator.

The ℓ^∞ norm

$$J_{H_\infty}(x) = \|x\|_\infty = \max_{1 \leqslant i \leqslant N} |x_i|$$

corresponds to choosing $H_\infty = [\mathrm{Id}_N, -\mathrm{Id}_N] \in \mathbb{R}^{N \times 2N}$. This regularization, coined anti-sparse regularization, is used for instance for approximate nearest neighbor search [3].

Another possible instance of polyhedral regularization is the group $\ell^1 - \ell^\infty$ regularization. Let $\mathcal{B}$ be a partition of $\{1, \ldots, N\}$. The $\ell^1 - \ell^\infty$ norm associated to this group structure is

$$J_{H_{\mathcal{B}}^\infty}(x) = \sum_{b \in \mathcal{B}} \|x_b\|_\infty.$$

This amounts to choosing the block-diagonal matrix $H_{\mathcal{B}}^\infty \in \mathbb{R}^{N \times \prod_{b \in \mathcal{B}} 2|b|}$ such that each column is chosen by taking for each block a position with sign ± 1, others are 0. If for all $b \in \mathcal{B}, |b| = 1$, then we recover the ℓ^1-norm, whereas if the block structure is composed by one element, we get the ℓ^∞-norm.

C. Prior Work

In the special case of ℓ^1 and analysis ℓ^1 penalties, our criterion is equivalent to those defined in [4] and [5]. To our knowledge, there is no generic guarantee for robustness to noise with ℓ^∞ regularization, but [6] studies robustness of a sub-class of polyhedral norms obtained by convex relaxation of combinatorial penalties. Its notion of support is however completely different from ours. The work [7] studies numerically some polyhedral regularizations.In [8], the authors provide an homotopy-like algorithm for polyhedral regularization through a continuous problem coined adaptive inverse scale space method. The work [9] analyzes some particular polyhedral regularizations in a noiseless compressed sensing setting when the matrix Φ is drawn from an appropriate random ensemble. Again in a compressed sensing scenario, the work of [10] studies a subset of polyhedral regularizations to get sharp estimates of the number of measurements for exact and ℓ_2-stable recovery.

II. Contributions

Definition 1. *We define the H-*support $\mathrm{supp}_H(x)$ *of a vector $x \in \mathbb{R}^N$ to be the set*

$$\mathrm{supp}_H(x) = \{i \in \{1,\dots,N_H\} \mid \langle x,\, h_i\rangle = J_H(x)\}\,.$$

This definition suggests that to recover signals with *H-support* $\mathrm{supp}_H(x)$, it would be reasonable to impose that Φ is invertible on the corresponding subspace $\mathrm{Ker}\, H^*_{\mathrm{supp}_H(x)}$. This is formalised in the following condition.

Definition 2. *A H-support I satisfies the* restricted injectivity condition *if*

$$\mathrm{Ker}\,\Phi \cap \mathrm{Ker}\, H_I^* = \{0\}, \qquad (\mathcal{C}_I)$$

where H_I is the matrix whose columns are those of H indexed by I.

When it holds, we define the orthogonal projection Γ_I on $\Phi\,\mathrm{Ker}\, H_I^*$:

$$M_I = (U^*\Phi^*\Phi U)^{-1} \quad \text{and} \quad \begin{cases}\Gamma_I &= \Phi U M_I U^* \Phi^* \\ \Gamma_I^\perp &= \mathrm{Id} - \Gamma_I.\end{cases}$$

where U is (any) basis of $\mathrm{Ker}\, H_I^*$. The symmetric bilinear form on $\mathbb{R}^N$ induced by $\Gamma_I^\perp$ reads

$$\langle u,\, v\rangle_{\Gamma_I^\perp} = \langle u,\, \Gamma_I^\perp v\rangle,$$

and we denote its associated quadratic form $\|\cdot\|^2_{\Gamma_I^\perp}$.

Definition 3. *Let I be a H-support such that $(\mathcal{C}_I)$ holds. The* Identifiability Criterion *of I is*

$$\mathbf{IC}_H(I) = \max_{z_I \in \mathrm{Ker}\, H_I} \min_{i \in I} (\tilde{\Phi}_I^* \Gamma_I^\perp \tilde{\Phi}_I \mathbb{I}_I + z_I)_i$$

where $\mathbb{I}_I \in \mathbb{R}^{|I|}$ is the vector with coefficients 1, and $\tilde{\Phi}_I = \Phi H_I^{+,*} \in \mathbb{R}^{Q\times|I|}$ where $^+$ stands for the Moore–Penrose pseudo-inverse.

$\mathbf{IC}_H(I)$ can be computed by solving the linear program

$$\mathbf{IC}_H(I) = \max_{(r,z_I)\in\mathbb{R}\times\mathbb{R}^{|I|}} r \text{ subj. to } \begin{cases}\forall i \in I, r \leqslant (\tilde{\Phi}_I^*\Gamma_I^\perp\tilde{\Phi}_I\mathbb{I}_I + z_I)_i \\ H_I z_I = 0.\end{cases}$$

A. Noise Robustness

Our main contribution is the following result.

Theorem 1. *Let $x_0 \in \mathbb{R}^N \setminus \{0\}$ and I its H-support such that $(\mathcal{C}_I)$ holds. Let $y = \Phi x_0 + w$. Suppose that $\tilde{\Phi}_I\mathbb{I}_I \neq 0$ and $\mathbf{IC}_H(I) > 0$. Then there exists two constants $c_I, \tilde{c}_I$ satisfying,*

$$\frac{\|w\|_2}{T} < \frac{\tilde{c}_I}{c_I} \quad \text{where} \quad T = \min_{j\in I^c} J_H(x_0) - \langle x_0,\, h_j\rangle > 0,$$

such that if λ is chosen according to

$$c_I\|w\|_2 < \lambda < T\tilde{c}_I,$$

the vector $x^\star \in \mathbb{R}^N$ defined by

$$x^\star = \mu H_I^{+,*}\mathbb{I}_I + U M_I U^* \Phi^*(y - \mu\tilde{\Phi}_I\mathbb{I}_I)$$

where U is any basis of $\mathrm{Ker}\, H_I^*$ and

$$0 < \mu = J_H(x_0) + \frac{\langle \tilde{\Phi}_I\mathbb{I}_I,\, w\rangle_{\Gamma_I^\perp} - \lambda}{\|\tilde{\Phi}_I\mathbb{I}_I\|^2_{\Gamma_I^\perp}} \qquad (2)$$

is the unique solution of $(P_\lambda(y))$, and $x^\star$ lives on the same face as x_0, i.e. $\mathrm{supp}_H(x^\star) = \mathrm{supp}_H(x_0)$.

Observe that if λ is chosen proportional to the noise level, then $\|x^\star - x_0\|_2 = O(\|w\|_2)$. The following proposition proves that the condition $\mathbf{IC}_H(I) > 0$ is almost a necessary condition to ensure the stability of the H-support. Its proof is omitted for obvious space limitation reasons.

Proposition 1. *Let $x_0 \in \mathbb{R}^N \setminus \{0\}$ and I its H-support such that $(\mathcal{C}_I)$ holds. Let $y = \Phi x_0 + w$. Suppose that $\tilde{\Phi}_I\mathbb{I}_I \neq 0$ and $\mathbf{IC}_H(I) < 0$. If $\frac{\|w\|}{\lambda} < \frac{1}{c_I}$ then for any solution of $(P_\lambda(y))$, we have $\mathrm{supp}_H(x_0) \neq \mathrm{supp}_H(x^\star)$.*

B. Noiseless Identifiability

When there is no noise, the following result, which is a straightforward consequence of Theorem 1, shows that the condition $\mathbf{IC}_H(I) > 0$ implies signal identifiability.

Theorem 2. *Let $x_0 \in \mathbb{R}^N \setminus \{0\}$ and I its H-support. Suppose that $\tilde{\Phi}_I\mathbb{I}_I \neq 0$ and $\mathbf{IC}_H(I) > 0$. Then the vector x_0 is the unique solution of $(P_0(y))$.*

III. Proofs

A. Preparatory Lemmata

We recall the definition of the subdifferential of a convex function f at the point x is the set $\partial f(x)$ is

$$\partial f(x) = \left\{ g \in \mathbb{R}^N \mid f(y) \geqslant f(x) + \langle g,\, y - x\rangle \right\}.$$

The following lemma, which is a direct consequence of the properties of the max function, gives the subdifferential of the regularization function J_H.

Lemma 1. *The subdifferential ∂J_H at $x \in \mathbb{R}^N$ reads*

$$\partial J_H(x) = H_I \Sigma_I$$

where $I = \mathrm{supp}_H(x)$ and Σ_I is the canonical simplex on $\mathbb{R}^{|I|}$:

$$\Sigma_I = \left\{ v_I \in \mathbb{R}^{|I|} \mid v_I \geqslant 0, \langle v_I,\, \mathbb{I}_I\rangle = 1 \right\}.$$

A point $x^\star$ is a minimizer of $\min_x f(x)$ if, and only if, $0 \in \partial f(x^\star)$. Thanks to Lemma 1, this gives the first-order condition for the problem $(P_\lambda(y))$.

Lemma 2. *A vector $x^\star$ is a solution of $(P_\lambda(y))$ if, and only if, there exists $v_I \in \Sigma_I$ such that*

$$\Phi^*(\Phi x - y) + \lambda H_I v_I = 0,$$

where $I = \mathrm{supp}_H(x)$.

We now introduce the following so-called source condition.
$(\mathbf{SC}_x)$**:** For $I = \mathrm{supp}_H(x)$, there exists η and $v_I \in \Sigma_I$ such that:

$$\Phi^*\eta = H_I v_I \in \partial J_H(x).$$

Under the source condition, a sufficient uniqueness condition can be derived when v_I lives in the relative interior of Σ_I which is

$$\mathrm{ri}\,\Sigma_I = \left\{ v_I \in \mathbb{R}^{|I|} \mid v_I > 0, \langle v_I,\, \mathbb{I}_I\rangle = 1 \right\}.$$

Lemma 3. *Let $x^\star$ be a minimizer of $(P_\lambda(y))$ (resp. $(P_0(y))$) and $I = \mathrm{supp}_H(x^\star)$. Assume that $(\mathbf{SC}_{x^\star})$ is verified with $v_I \in \mathrm{ri}\,\Sigma_I$, and that $(\mathcal{C}_I)$ holds. Then $x^\star$ is the unique solution of $(P_\lambda(y))$ (resp. $(P_0(y))$).*

The proof of this lemma is omitted due to lack of space. Observe that in the noiseless case, if the assumptions of Lemma 3 hold at x_0, then the latter is exactly recovered by solving $(P_0(y))$.

Lemma 4. *Let $x^\star \in \mathbb{R}^N$ and $I = \operatorname{supp}_H(x^\star)$. Assume $(\mathcal{C}_I)$ holds. Let U be any basis of $\operatorname{Ker} H_I^*$. There exists $z_I \in \operatorname{Ker} H_I$ such that*

$$U^*\Phi^*(\Phi x^\star - y) = 0$$
$$v_I = z_I + \frac{1}{\lambda} H_I^+ \Phi^*(y - \Phi x^\star) \in \Sigma_I,$$

if, and only if, $x^\star$ is a solution of $(P_\lambda(y))$. Moreover, if $v_I \in \operatorname{ri}\Sigma_I$, then $x^\star$ is the unique solution of $(P_\lambda(y))$.

Proof: We compute

$$\begin{aligned}&\Phi^*(\Phi x^\star - y) + \lambda H_I v_I\\ =&\Phi^*(\Phi x^\star - y) + \lambda H_I \left(z_I + \frac{1}{\lambda} H_I^+ \Phi^*(y - \Phi x^\star)\right)\\ =&(\mathrm{Id} - H_I H_I^+)\Phi^*(\Phi x^\star - y) = \operatorname{proj}_{H_I^*}(\Phi^*(\Phi x^\star - y)) = 0,\end{aligned}$$

where $\operatorname{proj}_{H_I^*}$ is the projection on $\operatorname{Ker} H_I^*$. Hence, $x^\star$ is a solution of $(P_\lambda(y))$. If $v_I \in \operatorname{ri}\Sigma_I$, then according to Lemma 3, $x^\star$ is the unique solution. ∎

The following lemma is a simplified rewriting of the condition introduced in Lemma 4.

Lemma 5. *Let $x^\star \in \mathbb{R}^N$, $I = \operatorname{supp}_H(x^\star)$ and $\mu = J_H(x^\star)$. Assume $(\mathcal{C}_I)$ holds. Let U be any basis of $\operatorname{Ker} H_I^*$. There exists $z \in \operatorname{Ker} H_I$ such that*

$$v_I = z_I + \frac{1}{\lambda}\tilde{\Phi}_I^* \Gamma_I^\perp (y - \mu\tilde{\Phi}_I \mathbb{I}_I) \in \Sigma_I,$$

if, and only if, $x^\star$ is a solution of $(P_\lambda(y))$. Moreover, if $v_I \in \operatorname{ri}\Sigma_I$, then $x^\star$ is the unique solution of $(P_\lambda(y))$.

Proof: Note that any vector $x \in \mathbb{R}^N$ such that the condition $(\mathcal{C}_I)$ holds, where I is the H-support of x, is such that

$$x = \mu H_I^{+,*}\mathbb{I}_I + U\alpha \quad \text{where} \quad \mu = J_H(x),$$

for some coefficients α and U any basis of $\operatorname{Ker} H_I^*$. We obtain

$$U\Phi^*(\Phi x^\star - y) = \mu U\Phi^*\Phi H_I^{+,*}\mathbb{I}_I - U\Phi^* y + U\Phi^*\Phi U\alpha = 0$$

Since $(\mathcal{C}_I)$ holds, we have

$$\alpha = (U\Phi^*\Phi U\alpha)^{-1} U\Phi^* \left(y - \mu\tilde{\Phi}_I \mathbb{I}_I\right).$$

Hence,

$$\Phi U\alpha = \Gamma_I \left(y - \mu\tilde{\Phi}_I\mathbb{I}_I\right).$$

Now since, $x^\star = \mu H_I^{+,*}\mathbb{I}_I + U\alpha$, one has

$$\Phi x^\star = \mu\tilde{\Phi}_I\mathbb{I}_I + \Gamma_I\left(y - \mu\tilde{\Phi}_I\mathbb{I}_I\right) = \mu\Gamma_I^\perp\tilde{\Phi}_I\mathbb{I}_I + \Gamma_I y.$$

Subtracting y and multiplying by $\tilde{\Phi}_I^*$ both sides, and replacing in the expression of v_I in Lemma 4, we get the desired result. ∎

B. Proof of Theorem 1

Let I be the H-support of x_0. We consider the restriction of $(P_\lambda(y))$ to the H-support I.

$$x^\star = \underset{\substack{x\in\mathbb{R}^N\\ \operatorname{supp}_H(x)\subseteq I}}{\operatorname{argmax}} \ \frac{1}{2}\|y - \Phi x\|_2^2 + J_H(x). \qquad (\mathcal{P}_\lambda(y)_I)$$

Thanks to $(\mathcal{C}_I)$, the objective function is strongly convex on the set of signals of H-support I. Hence $x^\star$ is uniquely defined. The proof is divided in five parts: We give (**1.**) an implicit form of $x^\star$. We check (**2.**) that the H-support of $x^\star$ is the same as the H-support of x_0. We provide (**3.**) the value of $J_H(x^\star)$. Using Lemma 5, we prove (**4.**) that $x^\star$ is the unique minimizer of $(P_\lambda(y))$.

1. Expression of $x^\star$. One has $x^\star = \mu H_I^{+,*}\mathbb{I}_I + U\alpha$ where $\mu = J_H(x^\star)$. Hence,

$$U^*\Phi^*(\Phi x - y) = \mu U^*\Phi^*\Phi H_I^{+,*}\mathbb{I}_I + (U^*\Phi^*\Phi U)\alpha - U^*\Phi^* y = 0.$$

Thus,

$$U\alpha = UM_I U^*\Phi^*(y - \mu\Phi H_I^{+,*}\mathbb{I}_I).$$

Now, since $y = \Phi x_0 + w$, with $\operatorname{supp}_H(x_0) = I$, then

$$\begin{aligned}x^\star &= \mu H_I^{+,*}\mathbb{I}_I + UM_I U^*\Phi^*(y - \mu\Phi H_I^{+,*}\mathbb{I}_I)\\ &= \mu H_I^{+,*}\mathbb{I}_I + UM_I U^*\Phi^*((\mu_0 - \mu)\Phi H_I^{+,*}\mathbb{I}_I + w) + U\alpha_0\\ &= x_0 - (\mu_0 - \mu)H_I^{+,*}\mathbb{I}_I + UM_I U^*\Phi^*((\mu_0 - \mu)\Phi H_I^{+,*}\mathbb{I}_I + w),\end{aligned}$$

where $\mu_0 = J_H(x_0)$. Hence, $x^\star$ is satisfying

$$x^\star = x_0 + (\mu_0 - \mu)[UM_I U^*\Phi^*\Phi - \mathrm{Id}]H_I^{+,*}\mathbb{I}_I + UM_I U^*\Phi^* w. \quad (3)$$

2. Checking that the H-support of $x^\star$ is I. To ensure that the H-support of $x^\star$ is I we have to impose that

$$\begin{aligned}\forall i \in I, &\quad \langle h_i,\, x^\star\rangle = J_H(x^\star) = \mu\\ \forall j \in I^c, &\quad \langle h_j,\, x^\star\rangle < J_H(x^\star) = \mu.\end{aligned}$$

The components on I of $x^\star$ are satisfying $H_I^* x^\star = \mu\mathbb{I}_I$. Since J_H is subadditive, we bound the components on I^c by the triangular inequality on (3) to get

$$\begin{aligned}\max_{j\in I^c}\langle h_j,\, x^\star\rangle \leqslant &\max_{j\in I^c}\langle h_j,\, x_0\rangle\\ &+ (\mu_0 - \mu)\|H_{I^c}^*[UM_I U^*\Phi^*\Phi - \mathrm{Id}]H_I^{+,*}\mathbb{I}_I\|_\infty\\ &+ \|H_{I^c}^* UM_I U^*\Phi^* w\|_\infty.\end{aligned}$$

Denoting

$$\begin{aligned}C_1 &= \|H_{I^c}^*[UM_I U^*\Phi^*\Phi - \mathrm{Id}]H_I^{+,*}\mathbb{I}_I\|_\infty,\\ C_2 &= \|H_{I^c}^* UM_I U^*\Phi^*\|_{2,\infty},\\ T &= \mu_0 - \max_{j\in I^c}\langle h_j,\, x_0\rangle,\end{aligned}$$

we bound the correlations outside the H-support by

$$\max_{j\in I^c}\langle h_j,\, x^\star\rangle \leqslant \mu_0 - T + (\mu_0 - \mu)C_1 + C_2\|w\|.$$

There exists some constants c_1, c_2 satisfying $c_1\|w\| < c_2 T + \lambda$ such that

$$0 \leqslant \mu_0 - T + (\mu_0 - \mu)C_1 + C_2\|w\| < \mu \quad (4)$$

Under this condition, one has

$$\max_{j\in I^c}\langle h_j,\, x^\star\rangle < \mu,$$

which proves that $\operatorname{supp}_H(x^\star) = I$.

3. Value of $\mu = J_H(x^\star)$. Using Lemma 5 with $H = U^*H$, since $x^\star$ is a solution of $(\mathcal{P}_\lambda(y)_I)$, there exists $z_I \in \operatorname{Ker} H_I$ such that

$$v_I = z_I + \frac{1}{\lambda}\tilde{\Phi}_I^*\Gamma_I^\perp(y - \mu\tilde{\Phi}_I\mathbb{I}_I) \in \Sigma_I. \quad (5)$$

We decompose x_0 as

$$x_0 = \mu_0 H_I^{+,*}\mathbb{I}_I + U\alpha_0.$$

Since $y = \Phi x_0 + w$, we have

$$\Gamma_I^\perp y = \Gamma_I^\perp(\mu_0\tilde{\Phi}_I\mathbb{I}_I + \Phi U\alpha_0 + w).$$

Now since

$$\Gamma_I\Phi U\alpha_0 = \Phi U(U^*\Phi^*\Phi U)^{-1}U^*\Phi^*\Phi U\alpha_0 = \Phi U\alpha_0,$$

one obtains

$$\Gamma_I^\perp y = \mu_0 \Gamma_I^\perp \tilde{\Phi}_I \mathbb{I}_I + \Gamma_I^\perp w.$$

Thus, equation (5) equivalently reads

$$v_I = z_I + \frac{1}{\lambda} \tilde{\Phi}_I^* \Gamma_I^\perp \left((\mu_0 - \mu) \tilde{\Phi}_I \mathbb{I}_I + w \right).$$

In particular, $\langle v_I, \mathbb{I}_I \rangle = \lambda$. Thus,

$$\lambda = \langle \lambda v_I, \mathbb{I}_I \rangle = \langle \lambda \tilde{z}_I, \mathbb{I}_I \rangle + \langle \tilde{\Phi}_I^* \Gamma_I^\perp ((\mu_0 - \mu) \tilde{\Phi}_I \mathbb{I}_I + w, \mathbb{I}_I \rangle.$$

Since $\tilde{z}_I \in \operatorname{Ker} H_I$, one has $\langle z_I, \mathbb{I}_I \rangle = 0$.

$$\begin{aligned} \lambda &= \langle \tilde{\Phi}_I^* \Gamma_I^\perp ((\mu_0 - \mu) \tilde{\Phi}_I \mathbb{I}_I + w, \mathbb{I}_I \rangle \\ &= (\mu_0 - \mu) \|\tilde{\Phi}_I \mathbb{I}_I\|_{\Gamma_I^\perp}^2 + \langle \tilde{\Phi}_I \mathbb{I}_I, w \rangle_{\Gamma_I^\perp}. \end{aligned}$$

Thus the value of μ is given by

$$\mu = \mu_0 + \frac{\langle \tilde{\Phi}_I \mathbb{I}_I, w \rangle_{\Gamma_I^\perp} - \lambda}{\|\tilde{\Phi}_I \mathbb{I}_I\|_{\Gamma_I^\perp}^2} > 0. \tag{6}$$

4. Checking conditions of Lemma 5. Consider now the vector $\tilde{v}_I$ defined by

$$\tilde{v}_I = \tilde{z}_I + \frac{1}{\lambda} \tilde{\Phi}_I^* \Gamma_I^\perp \left((\mu_0 - \mu) \tilde{\Phi}_I \mathbb{I}_I + w \right),$$

where

$$\tilde{z}_I = \frac{1}{\mu - \mu_0} \left(\underset{z_I \in \operatorname{Ker} H_I}{\operatorname{argmax}} \ \min_{i \in I} (\tilde{\Phi}_I^* \Gamma_I^\perp \tilde{\Phi}_I \mathbb{I}_I + z_I)_i \right)$$

Under condition (4), the H-support of $x^\star$ is I, hence we only have to check that $\tilde{v}_I$ is an element of $\operatorname{ri} \Sigma_I$. Since $\langle \tilde{z}_I, \mathbb{I}_I \rangle = 0$, one has

$$\begin{aligned} &\langle \tilde{v}_I, \mathbb{I}_I \rangle \\ =&\langle z_I + \frac{1}{\lambda} \tilde{\Phi}_I^* \Gamma_I^\perp \left((\mu_0 - \mu) \tilde{\Phi}_I \mathbb{I}_I + w \right), \mathbb{I}_I \rangle + \langle \tilde{z}_I - z_I, \mathbb{I}_I \rangle \\ =&\langle v_I, \mathbb{I}_I \rangle + 0 = \lambda. \end{aligned}$$

Plugging back the expression (6) of $(\mu_0 - \mu)$ in the definition of $\tilde{v}_I$, one has

$$\tilde{v}_I = \tilde{z}_I + \frac{1}{\lambda} \left(\tilde{\Phi}_I^* \Gamma_I^\perp w + \frac{\langle \tilde{\Phi}_I \mathbb{I}_I, w \rangle_{\Gamma_I^\perp} - \lambda}{\|\tilde{\Phi}_I \mathbb{I}_I\|_{\Gamma_I^\perp}^2} \tilde{\Phi}_I^* \Gamma_I^\perp \tilde{\Phi}_I \mathbb{I}_I \right).$$

For some constant c_3 such that $c_3 \|w\| - \mathbf{IC}_H(I) \cdot \lambda > 0$, one has

$$\forall i \in I, \quad v_i > 0.$$

Combining this with the fact that $\langle \tilde{v}_I, \mathbb{I}_I \rangle = \lambda$ proves that $\tilde{v}_I \in \operatorname{ri} \Sigma_I$. According to Lemma 5, $x^\star$ is the unique minimizer of $(P_\lambda(y))$. ■

C. Proof of Theorem 2

Taking $w = 0$ in Theorem 1, we obtain immediately

Lemma 6. *Let $x_0 \in \mathbb{R}^N \setminus \{0\}$ and I its H-support such that $(\mathcal{C}_I)$ holds. Let $y = \Phi x_0$. Suppose that $\tilde{\Phi}_I \mathbb{I}_I \neq 0$ and $\mathbf{IC}_H(I) > 0$. Let $T = \min_{j \in I^c} J_H(x_0) - \langle x_0, h_j \rangle > 0$ and $\lambda < T \tilde{c}_I$. Then,*

$$x^\star = x_0 + \frac{\lambda}{\|\tilde{\Phi}_I \mathbb{I}_I\|_{\Gamma_I^\perp}^2} [U M_I U^* \Phi^* \Phi - \operatorname{Id}] H_I^{+,*} \mathbb{I}_I,$$

is the unique solution of $(P_\lambda(y))$.

The following lemma shows that under the same condition, x_0 is a solution of $(P_0(y))$.

Lemma 7. *Let $x_0 \in \mathbb{R}^N \setminus \{0\}$ and I its H-support such that $(\mathcal{C}_I)$ holds. Let $y = \Phi x_0$. Suppose that $\tilde{\Phi}_I \mathbb{I}_I \neq 0$ and $\mathbf{IC}_H(I) > 0$. Then x_0 is a solution of $(P_0(y))$.*

Proof: According to Lemma 6, for every $0 < \lambda < T \tilde{c}_I$,

$$x_\lambda^\star = x_0 + \frac{\lambda}{\|\tilde{\Phi}_I \mathbb{I}_I\|_{\Gamma_I^\perp}^2} [U M_I U^* \Phi^* \Phi - \operatorname{Id}] H_I^{+,*} \mathbb{I}_I,$$

is the unique solution of $(P_\lambda(y))$.

Let $\tilde{x} \neq x_0$ such that $\Phi \tilde{x} = y$. For every $0 < \lambda < T \tilde{c}_I$, since $x_\lambda^\star$ is the unique minimizer of $(P_\lambda(y))$, one has

$$\frac{1}{2} \|y - \Phi x_\lambda^\star\|_2^2 + J_H(x_\lambda^\star) < \frac{1}{2} \|y - \Phi \tilde{x}\|_2^2 + J_H(\tilde{x}).$$

Using the fact that $\Phi \tilde{x} = y = \Phi x_0$, one has $J_H(x_\lambda^\star) < J_H(\tilde{x})$. By continuity of the mapping $x \mapsto J_H(x)$, taking the limit for $\lambda \to 0$ in the previous inequality gives

$$J_H(x_0) \leqslant J_H(\tilde{x}).$$

It follows that x_0 is a solution of $(P_0(y))$. ■

We now prove Theorem 2.

Proof of Theorem 2: Lemma 7 proves that x_0 is a solution of $(P_0(y))$. We now prove that x_0 is in fact the unique solution. Let $\tilde{z}_I$ be the argument of the maximum in the definition of $\mathbf{IC}_H(I)$. We define

$$\tilde{v}_I = \frac{1}{\|\tilde{\Phi}_I \mathbb{I}_I\|_{\Gamma_I^\perp}^2} \left(\tilde{z}_I + \tilde{\Phi}_I^* \Gamma_I^\perp \tilde{\Phi}_I \mathbb{I}_I \right).$$

By definition of $\mathbf{IC}_H(I)$, for every $i \in I, \tilde{v}_I > 0$ and $\langle \tilde{v}_I, \mathbb{I}_I \rangle = 1$. Thus, $H_I \tilde{v}_I \in \operatorname{ri}(\partial J_H(x_0))$. Moreover, since $\tilde{z}_I \in \operatorname{Ker} H_I$, one has

$$H_I v_I = H_I H_I^{+,*} \Phi^* \Gamma_I^\perp \tilde{\Phi}_I \mathbb{I}_I = \Phi^* \eta \quad \text{where} \quad \eta = \Gamma_I^\perp \tilde{\Phi}_I \mathbb{I}_I.$$

Thanks to Lemma 3, x_0 is the unique solution of $(P_0(y))$. ■

References

[1] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[2] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[3] H. Jégou, T. Furon, and J. Fuchs, "Anti-sparse coding for approximate nearest neighbor search," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2029–2032.

[4] J. Fuchs, "On sparse representations in arbitrary redundant bases," *Information Theory, IEEE Transactions on*, vol. 50, no. 6, pp. 1341–1344, 2004.

[5] S. Vaiter, G. Peyré, C. Dossal, and J. Fadili, "Robust sparse analysis regularization," *to appear in IEEE Transactions on Information Theory*, 2012.

[6] F. Bach, "Structured sparsity-inducing norms through submodular functions," *Advances in Neural Information Processing Systems*, 2010.

[7] S. Petry and G. Tutz, "Shrinkage and variable selection by polytopes," *Journal of Statistical Planning and Inference*, vol. 142, no. 1, pp. 48–64, 2012.

[8] M. Moeller and M. Burger, "Multiscale methods for polyhedral regularizations," UCLA, CAM Report 11-74, 2011.

[9] D. Donoho and J. Tanner, "Counting the faces of randomly-projected hypercubes and orthants, with applications," *Discrete & computational geometry*, vol. 43, no. 3, pp. 522–541, 2010.

[10] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, "The Convex Geometry of Linear Inverse Problems," *Foundations of Computational Mathematics*, vol. 12, no. 6, pp. 805–849, 2012.