Robust Polyhedral Regularization
Samuel Vaiter, Gabriel Peyré, Jalal M. Fadili

To cite this version:
Samuel Vaiter, Gabriel Peyré, Jalal M. Fadili. Robust Polyhedral Regularization. International Conference on Sampling Theory and Applications (SampTA), 2013, Bremen, Germany. hal-00816377

HAL Id: hal-00816377
https://hal.archives-ouvertes.fr/hal-00816377
Submitted on 22 Apr 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Robust Polyhedral Regularization

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 $\ell_2$ 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 $\ell^1$ regularization which are encompassed in this framework. As a byproduct, we obtain recovery guarantees for $\ell^\infty$ and $\ell^1 - \ell^\infty$ regularization.

I. INTRODUCTION

A. Polyhedral Regularization

We consider the following linear inverse problem

$$y = \Phi x_0 + w,$$

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 $\Phi$ 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 $P$ of $\mathbb{R}^N$ such that $P = \{x \in \mathbb{R}^N | Ax \leq b\}$ for some $A \in \mathbb{R}^{N \times H} \times N$ and $b \in \mathbb{R}^N$. The inequality $\leq$ 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

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

where $H = (h_i)_{i=1}^N \in \mathbb{R}^{N \times N}$. Thus, $P_H = \{x \in \mathbb{R}^N | J_H(x) \leq 1\}$ is a polyhedron. We assume that $P_H$ 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

$$x^* \in \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \|y - \Phi x\|^2 + \lambda J_H(x), \quad (P_\lambda(y))$$

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

$$x^* \in \arg\min_{x \in y} J_H(x), \quad (P_0(y))$$

$B$. Relation to Sparsity and Anti-sparsity

Examples of polyhedral regularization include the $\ell^1$-norm, analysis $\ell^1$-norm and $\ell^\infty$-norm. The $\ell^1$ norm reads

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

It corresponds to choosing $H_1 \in \mathbb{R}^{N \times 2N}$ 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 the true vector is equal to that of the recovered one. This criterion also applies to noise perturbations of polyhedral regularization.

C. Prior Work

In the special case of $\ell^1$ and analysis $\ell^1$ 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 $\ell^\infty$ 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 $\Phi$ is drawn from an appropriate random ensemble. Again in a compressed sensing scenario, the work of [10] studies a subset of polyhedral

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 $\ell_2$ 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 $\ell^1$ regularization which are encompassed in this framework. As a byproduct, we obtain recovery guarantees for $\ell^\infty$ and $\ell^1 - \ell^\infty$ regularization.

I. INTRODUCTION

A. Polyhedral Regularization

We consider the following linear inverse problem

$$y = \Phi x_0 + w,$$

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 $\Phi$ 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 $P$ of $\mathbb{R}^N$ such that $P = \{x \in \mathbb{R}^N | Ax \leq b\}$ for some $A \in \mathbb{R}^{N \times H} \times N$ and $b \in \mathbb{R}^N$. The inequality $\leq$ 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

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

where $H = (h_i)_{i=1}^N \in \mathbb{R}^{N \times N}$. Thus, $P_H = \{x \in \mathbb{R}^N | J_H(x) \leq 1\}$ is a polyhedron. We assume that $P_H$ 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

$$x^* \in \arg\min_{x \in \mathbb{R}^N} \frac{1}{2} \|y - \Phi x\|^2 + \lambda J_H(x), \quad (P_\lambda(y))$$

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

$$x^* \in \arg\min_{x \in y} J_H(x), \quad (P_0(y))$$

B. Relation to Sparsity and Anti-sparsity

Examples of polyhedral regularization include the $\ell^1$-norm, analysis $\ell^1$-norm and $\ell^\infty$-norm. The $\ell^1$ norm reads

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

It corresponds to choosing $H_1 \in \mathbb{R}^{N \times 2N}$ 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 the true vector is equal to that of the recovered one. This criterion also applies to noise perturbations of polyhedral regularization.

C. Prior Work

In the special case of $\ell^1$ and analysis $\ell^1$ 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 $\ell^\infty$ 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 $\Phi$ 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 $\ell_2$-stable recovery.

II. Contributions

Definition 1. We define the $H$-support $\text{supp}_H(x)$ of a vector $x \in \mathbb{R}^N$ to be the set
$$\text{supp}_H(x) = \{ i \in \{1, \ldots, N_H\} \mid \langle x, h_i \rangle = J_H(x) \}.$$  

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

Definition 2. A $H$-support $I$ satisfies the restricted injectivity condition if
$$\text{Ker} \Phi \cap \text{Ker} H^*_I = \{0\}, \quad (C_I)$$
where $H_I$ is the matrix whose columns are those of $H$ indexed by $I$.

When it holds, we define the orthogonal projection $\Gamma_I$ on $\text{Ker} H^*_I$:
$$M_I = (U^* \Phi^* \Phi U)^{-1} \quad \text{and} \quad \Gamma_I = \Phi U M_I U^* \Phi^*$$
where $U$ is (any) basis of $\text{Ker} H^*_I$. The symmetric bilinear form on $\mathbb{R}^N$ induced by $\Gamma_I$ reads
$$\langle u, v \rangle_{\Gamma_I} = \langle u, \Gamma_I v \rangle,$$
and we denote its associated quadratic form $\| \cdot \|_{\Gamma_I^2}$.

Definition 3. Let $I$ be a $H$-support such that $(C_I)$ holds. The Identifiability Criterion of $I$ is
$$IC_H(I) = \max_{\xi \in \text{supp}_H(x)} \min_{\vartheta \in I} \langle \hat{\Phi}^*_I \Gamma_I^* \hat{\Phi}^*_I + z_i \rangle,$$
where $\xi \in \mathbb{R}^{|I|}$ is the vector with coefficients $I$, and $\hat{\Phi}^*_I = \Phi H^*_I \ast \in \mathbb{R}^{|Q| \times |I|}$ where $+$ stands for the Moore-Penrose pseudo-inverse.

$IC_H(I)$ can be computed by solving the linear program
$$IC_H(I) = \max_{(r,z) \in \mathbb{R}^{|I|} \times \mathbb{R}^{|I|}} \text{subj. to } \{ \forall i \in I, r \leq \langle \hat{\Phi}^*_I \Gamma_I^* \hat{\Phi}^*_I + z_i \rangle$$
$$\text{H}_I z_I = 0 \}.$$

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 $(C_I)$ holds. Let $y = \Phi x_0 + w$. Suppose that $\hat{\Phi}^*_I z_I \neq 0$ and $IC_H(I) > 0$. Then there exists two constants $c_I, \tilde{c}_I$ satisfying
$$\frac{|w|_2}{T} < \frac{\tilde{c}_I}{c_I} \quad \text{where} \quad T = \min_{j \in I} J^*_H(x_0) - (x_0, h_j) > 0,$$
such that if $\lambda$ is chosen according to
$$c_I |w|_2 < \lambda < T \tilde{c}_I,$$
the vector $x^* \in \mathbb{R}^N$ defined by
$$x^* = \mu H^*_I \ast \tilde{I}_I + UM_I U^* \Phi^* (y - \mu \hat{\Phi}^*_I \tilde{I}_I)$$
where $U$ is any basis of $\text{Ker} H^*_I$ and
$$0 < \mu = J_H(x_0) + \frac{\langle \hat{\Phi}^*_I \tilde{I}_I, w \rangle_{\Gamma_I^*} - \lambda}{\| \hat{\Phi}^*_I \tilde{I}_I \|_{\Gamma_I^2}} \quad (2)$$
is the unique solution of $\langle P_\lambda(y) \rangle$, and $x^*$ lives on the same face as $x_0$, i.e. $\text{supp}_H(x^*) = \text{supp}_H(x_0)$.

Observe that if $\lambda$ is chosen proportional to the noise level, then $|x^* - x_0|_2 = O(|w|_2)$. The following proposition proves that the condition $IC_H(I) > 0$ 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 $(C_I)$ holds. Let $y = \Phi x_0 + w$. Suppose that $\hat{\Phi}^*_I z_I \neq 0$ and $IC_H(I) < 0$. If
$$\frac{|w|_2}{T} < \frac{1}{\tilde{c}_I}$$
then for any solution of $\langle P_\lambda(y) \rangle$, we have $\text{supp}_H(x_0) \neq \text{supp}_H(x^*)$.

B. Noiseless Identifiability

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

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

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

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

Lemma 1. The subdifferential $\partial J_H(x)$ at $x \in \mathbb{R}^N$ reads
$$\partial J_H(x) = H_I \Sigma_I$$
where $I = \text{supp}_H(x)$ and $\Sigma_I$ is the canonical simplex on $\mathbb{R}^{|I|}$:
$$\Sigma_I = \{ v_I \in \mathbb{R}^{|I|} \mid v_I \geq 0, \{ v_I, \xi_I = 1 \} \}.$$  

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

Lemma 2. A vector $x^*$ is a solution of $\langle P_\lambda(y) \rangle$ if, and only if, there exists $v_I \in \Sigma_I$ such that
$$\Phi^* (\Phi x - y) + \lambda H_I v_I = 0,$$
where $I = \text{supp}_H(x)$.

We now introduce the following so-called source condition.

$(SC_{\mu})$: For $I = \text{supp}_H(x)$, there exists $\eta$ 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 $\Sigma_I$ which is
$$\{ v_I \in \mathbb{R}^{|I|} \mid v_I > 0, \{ v_I, \xi_I = 1 \} \}.$$  

Lemma 3. Let $x^*$ be a minimizer of $\langle P_\lambda(y) \rangle$ (resp. $\langle P_0(y) \rangle$) and $I = \text{supp}_H(x^*)$. Assume that $(SC_{\mu^*})$ is verified with $v_I \in \Sigma_I$, and that $(C_I)$ holds. Then $x^*$ is the unique solution of $\langle P_\lambda(y) \rangle$ (resp. $\langle P_0(y) \rangle$).

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 $\langle P_0(y) \rangle$. 

Lemma 4. Let \( x^* \in \mathbb{R}^N \) and \( I = \text{supp}_H(x^*) \). Assume (C1) holds. Let \( U \) be any basis of \( \ker H_I^* \). There exists \( z_I \in \ker H_I \) such that
\[
V^* \Phi^* (\Phi x^* - y) = 0
\]
if, and only if, \( x^* \) is a solution of \((P_\lambda(y))\). Moreover, if \( v_I \in \mathbb{R}^{\Sigma_I} \), then \( x^* \) is the unique solution of \((P_\lambda(y))\).

Proof: We compute
\[
\Phi^* (\Phi x^* - y) + \lambda H_I v_I
= \Phi^* (\Phi x^* - y) + H_I \left( z_I + \frac{1}{\lambda} H_I^* (y - \Phi x^*) \right)
= (\text{Id} - H_I H_I^*) \Phi^* (\Phi x^* - y) = \text{proj}_{H_I^*} (\Phi^* (\Phi x^* - y)) = 0,
\]
where \( \text{proj}_{H_I^*} \) is the projection on \( \ker H_I^* \). Hence, \( x^* \) is a solution of \((P_\lambda(y))\). If \( v_I \in \mathbb{R}^{\Sigma_I} \), then according to Lemma 3, \( x^* \) is the unique solution.

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

Lemma 5. Let \( x^* \in \mathbb{R}^N \), \( I = \text{supp}_H(x^*) \) and \( \mu = J_H(x^*) \). Assume (C1) holds. Let \( U \) be any basis of \( \ker H_I^* \). There exists \( z \in \ker H_I \) such that
\[
v_I = z_I + \frac{1}{\lambda} \Phi_I^* \Gamma_I^* (y - \Phi_I z_I) \in \Sigma_I,
\]
if, and only if, \( x^* \) is a solution of \((P_\lambda(y))\). Moreover, if \( v_I \in \mathbb{R}^{\Sigma_I} \), then \( x^* \) is the unique solution of \((P_\lambda(y))\).

Proof: Note that any vector \( x \in \mathbb{R}^N \) such that the condition (C1) holds, where \( I \) is the \( H \)-support of \( x \), is such that
\[
x = \mu H_I^* \Xi_I + U \alpha \quad \text{where} \quad \mu = J_H(x),
\]
for some coefficients \( \alpha \) and \( U \) any basis of \( \ker H_I^* \). We obtain
\[
U \Phi^* (\Phi x^* - y) = \mu U \Phi^* \Phi H_I^* \Xi_I + U \Phi^* y + U \Phi^* U \alpha = 0
\]
Since (C1) holds, we have
\[
\alpha = (U \Phi^* U \alpha)^{-1} U \Phi^* \left( y - \mu \Phi_I \Xi_I \right).
\]

Hence,
\[
\Phi U \alpha = \Gamma_I \left( y - \mu \Phi_I \Xi_I \right).
\]
Now since, \( x^* = \mu \Phi_I z_I + \Gamma_I \left( y - \mu \Phi_I \Xi_I \right) = \mu \Gamma_I \Phi_I z_I + \Gamma_I y \).

Subtracting \( y \) and multiplying by \( \Phi_I^* \) 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 \( I \).
\[
x^* = \arg \max_{z \in \mathbb{R}^N} \frac{1}{2} \| y - \Phi x \|^2 + J_H(x) \quad (P_\lambda(y))
\]
Thanks to (C1), the objective function is strongly convex on the set of signals of \( H \)-support \( I \). Hence \( x^* \) is uniquely defined. The proof is divided in five parts: We give (1.) an implicit form of \( x^* \). We check (2.) that the \( H \)-support of \( x^* \) is the same as the \( H \)-support of \( x_0 \). We provide (3.) the value of \( J_H(x^*) \). Using Lemma 5, we prove (4.) that \( x^* \) is the unique minimizer of \((P_\lambda(y))\).

1. Expression of \( x^* \). One has \( x^* = \mu H_I^* \Xi_I + U \alpha \) where \( \mu = J_H(x^*) \).

Hence,
\[
U \Phi^* (\Phi x - y) = \mu U \Phi^* \Phi H_I^* \Xi_I + (U \Phi^* U \alpha - U \Phi^* y) = 0.
\]
The cost function is strongly convex on the set \( H(I) \). Now, since \( y = \Phi x_0 + w \), with \( \text{supp}_H(x_0) = I \), then
\[
x^* = \mu H_I^* \Xi_I + U M_I U^* \Phi^* (y - \mu \Phi_I \Xi_I + w) = U \alpha_0 + (\mu_0 - \mu) H_I^* \Xi_I + U M_I U^* \Phi^* (\mu_0 - \mu) \Phi_I \Xi_I + w,
\]
where \( \mu_0 = J_H(x_0) \). Hence, \( x^* \) is satisfying
\[
x^* = x_0 + (\mu_0 - \mu) [U M_I U^* \Phi^* - \text{Id}] H_I^* \Xi_I + U M_I U^* \Phi^* w.
\]

2. Checking that the \( H \)-support of \( x^* \) is \( I \). To ensure that the \( H \)-support of \( x^* \) is \( I \) we have to impose that
\[
\forall i \in I, \quad (h_i, x^*) = J_H(x^*) = \mu
\]
\[
\forall j \in I^c, \quad (h_j, x^*) < J_H(x^*) = \mu.
\]
The components on \( I \) are satisfying \( H_I^* x^* = \mu \Xi_I \). Since \( J_H \) is subadditive, we bound the components on \( I^c \) by the triangular inequality on (3) to get
\[
\max_{j \in I^c} (h_j, x^*) \leq \max_{j \in I^c} (h_j, x_0) + (\mu_0 - \mu) \| H_I^* U M_I U^* \Phi^* - \text{Id} \| H_I^* \Xi_I + U M_I U^* \Phi^* w.
\]

Denoting
\[
C_1 = \| H_I^* U M_I U^* \Phi^* - \text{Id} \| H_I^* \Xi_I + U M_I U^* \Phi^* w,
\]
\[
C_2 = \| H_I^* U M_I U^* \Phi^* \|_2,
\]
\[
T = \mu_0 - \max_{j \in I^c} (h_j, x_0),
\]
we bound the correlations outside the \( H \)-support by
\[
\max_{j \in I^c} (h_j, x^*) \leq \mu_0 - T + (\mu_0 - \mu) C_1 + C_2 \| w \|.
\]

There exists some constants \( c_1, c_2 \) satisfying \( c_1 \| w \| < c_2 T + \lambda \) such that
\[
0 \leq \mu_0 - T + (\mu_0 - \mu) C_1 + C_2 \| w \| < \mu.
\]
Under this condition, one has
\[
\max_{j \in I^c} (h_j, x^*) < \mu,
\]
which proves that \( \text{supp}_H(x^*) = I \).

3. Value of \( \mu = J_H(x^*) \). Using Lemma 5 with \( H = U^* H \), since \( x^* \) is a solution of \((P_\lambda(y))\), there exists \( z_I \in \ker H_I \) such that
\[
v_I = z_I + \frac{1}{\lambda} \Phi_I^* \Gamma_I^* (y - \mu \Phi_I z_I) \in \Sigma_I.
\]
We decompose \( x_0 \) as
\[
x_0 = \mu_0 H_I^* \Xi_I + U \alpha_0.
\]

Since \( y = \Phi x_0 + w \), we have
\[
\Gamma_I^* y = \Gamma_I^* (\mu_0 \Phi_I z_I + \Phi U \alpha_0 + w).
\]
Now since
\[
\Gamma_I \Phi U \alpha_0 = \Phi U (U^* \Phi^* \Phi U)^{-1} U^* \Phi^* U \alpha_0 = \Phi U \alpha_0,
\]
Thus, equation (5) equivalently reads
\[ v_T = z_T + \frac{1}{\lambda} \Phi_T^\top \Gamma_T^\perp \left( (\mu_0 - \mu) \Phi_T I_T + w \right). \]
In particular, \( \langle v_T, I_T \rangle = \lambda \). Thus,
\[ \lambda = \langle \lambda x_T, I_T \rangle = \langle \lambda \tilde{x}_T, I_T \rangle + \langle \Phi_T^\top \Gamma_T \left( (\mu_0 - \mu) \Phi_T I_T + w, I_T \right) \]
Since \( \tilde{x}_T \in Ker H_I \), one has \( \langle z_T, I_T \rangle = 0 \).
\[ \lambda = \langle \Phi_T^\top \Gamma_T \left( (\mu_0 - \mu) \Phi_T I_T + w, I_T \right) \]
\[ = (\mu_0 - \mu) \langle \Phi_T I_T, I_T \rangle + \langle \Phi_T I_T, w \rangle I_T. \]
Thus the value of \( \mu \) is given by
\[ \mu = \mu_0 + \frac{\langle \Phi_T I_T, I_T \rangle + \langle \Phi_T I_T, w \rangle I_T}{\| \Phi_T I_T \|^2_{I_T}} > 0. \quad (6) \]

4. Checking conditions of Lemma 5. Consider now the vector \( \tilde{v}_T \) defined by
\[ \tilde{v}_T = z_T + \frac{1}{\mu - \mu_0} \left( \text{argmax}_{\tilde{z}_T \in Ker H_I, V \notin T} \min \langle \Phi_T^\top \Gamma_T \tilde{z}_T, z_T \rangle \right), \]
Under condition (4), the \( H \)-support of \( x^* \) is \( I \), hence we only have to check that \( \tilde{v}_T \) is an element of \( r_i \Sigma_T \). Since \( \langle \tilde{z}_T, I_T \rangle = 0 \), one has
\[ \langle \tilde{v}_T, I_T \rangle = \langle z_T + \frac{1}{\mu - \mu_0} \left( \text{argmax}_{\tilde{z}_T \in Ker H_I, V \notin T} \min \langle \Phi_T^\top \Gamma_T \tilde{z}_T, z_T \rangle \right), I_T \rangle = \langle v_T, I_T \rangle = 0. \]
Plugging back the expression (6) of \( (\mu_0 - \mu) \) in the definition of \( \tilde{v}_T \), one has
\[ \tilde{v}_T = z_T + \frac{1}{\mu} \left( \langle \Phi_T I_T, I_T \rangle + \langle \Phi_T I_T, w \rangle I_T - \frac{1}{\| \Phi_T I_T \|^2_{I_T}} \langle \Phi_T^\top \Gamma_T \tilde{v}_T, I_T \rangle \right). \]
For some constant \( c_3 \) such that \( c_3 \| w \| - \| I C_H(I) \cdot \lambda \| > 0 \), one has
\[ \forall i \in I, \quad v_i > 0. \]
Combining this with the fact that \( \langle \tilde{v}_T, I_T \rangle = \lambda \) proves that \( \tilde{v}_T \in r_i \Sigma_T \). According to Lemma 5, \( x^* \) is the unique minimizer of \( (P_3(y)) \).

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 \( (C_I) \) holds. Let \( y = \Phi x_0 \). Suppose that \( \Phi_T x_T \neq 0 \) and \( IC_H(I) > 0 \). Let \( T = \min \{ J_H(x_0) - \{ x_0, h_j \}, h_j > 0 \} > \lambda < T \varepsilon_T \). Then,
\[ x^* = x_0 + \frac{\lambda}{\| \Phi_T I_T \|^2_{I_T}} \left[ U M I_T U^\top \Phi U^\top - 1d \right] H_T^{\perp \perp} I_T, \]
is the unique solution of \( (P_3(y)) \).

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 \( (C_I) \) holds. Let \( y = \Phi x_0 \). Suppose that \( \Phi_T x_T \neq 0 \) and \( IC_H(I) > 0 \). Then \( x_0 \) is a solution of \( (P_0(y)) \).

Proof: According to Lemma 6, for every \( 0 < \lambda < T \varepsilon_T \),
\[ x_0 = x_0 + \frac{\lambda}{\| \Phi_T I_T \|^2_{I_T}} \left[ U M I_T U^\top \Phi U^\top - 1d \right] H_T^{\perp \perp} I_T, \]
is the unique solution of \( (P_3(y)) \).

Let \( x_0 \neq x_0 \) such that \( \Phi x_0 = y \). For every \( 0 < \lambda < T \varepsilon_T \), since \( x_0 \) is the unique minimizer of \( (P_3(y)) \), one has
\[ \frac{1}{2} \| y - \Phi x_0 \|^2 + J_H(x_0) < \frac{1}{2} \| y - \Phi x_0 \|^2 + J_H(\tilde{x}). \]
Using the fact that \( \Phi \tilde{x} = y = \Phi x_0 \), one has \( J_H(x_0) < J_H(\tilde{x}) \). By continuity of the mapping \( x \mapsto J_H(x) \), taking the limit for \( \lambda \rightarrow 0 \) in the previous inequality gives
\[ J_H(x_0) \leq J_H(\tilde{x}). \]
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 \( \tilde{v}_T \) be the argument of the maximum in the definition of \( IC_H(I) \). We define
\[ \tilde{v}_T = \frac{1}{\| \Phi_T I_T \|^2_{I_T}} \left( z_T + \frac{\lambda}{\| \Phi_T I_T \|^2_{I_T}} \right), \]
By definition of \( IC_H(I) \), for every \( i \in I, \tilde{v}_T > 0 \) and \( \langle \tilde{v}_T, I_T \rangle = 1 \). Thus, \( H_T \tilde{v}_T \in r_i J_H(x_0) \). Moreover, since \( \tilde{z}_T \in Ker H_I \), one has
\[ H_T \tilde{v}_T = H_T H_T^{\perp \perp} \Phi^\top - \Phi^\top \tilde{v}_T = \Phi^\top \eta \quad \text{where} \quad \eta = \Gamma_T^\top \tilde{v}_T. \]
Thanks to Lemma 3, \( x_0 \) is the unique solution of \( (P_0(y)) \).

References