



HAL
open science

Model Selection with Piecewise Regular Gauges

Samuel Vaïter, Mohammad Golbabaee, Jalal M. Fadili, Gabriel Peyré

► **To cite this version:**

Samuel Vaïter, Mohammad Golbabaee, Jalal M. Fadili, Gabriel Peyré. Model Selection with Piecewise Regular Gauges. 2013. hal-00842603v1

HAL Id: hal-00842603

<https://hal.science/hal-00842603v1>

Submitted on 8 Jul 2013 (v1), last revised 20 May 2015 (v3)

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.

Model Selection with Piecewise Regular Gauges

SAMUEL VAITER¹, MOHAMMAD GOLBABAEE¹,
JALAL FADILI² AND GABRIEL PEYRÉ¹

¹*CNRS, CEREMADE, Université Paris-Dauphine,
{samuel.vaiter, golbabaee, peyre}@ceremade.dauphine.fr*
²*GREYC, CNRS-ENSICAEN-Université de Caen, jalal.fadili@ensicaen.fr*

Abstract

Regularization plays a pivotal role when facing the challenge of solving ill-posed inverse problems, where the number of observations is smaller than the ambient dimension of the object to be estimated. A line of recent work has studied regularization models with various types of low-dimensional structures. In such settings, the general approach is to solve a regularized optimization problem, which combines a data fidelity term and some regularization penalty that promotes the assumed low-dimensional/simple structure. This paper provides a general framework to capture this low-dimensional structure through what we coin piecewise regular gauges. These are convex, non-negative, closed, bounded and positively homogenous functions that will promote objects living on low-dimensional subspaces. This class of regularizers encompasses many popular examples such as the ℓ^1 norm, $\ell^1 - \ell^2$ norm (group sparsity), nuclear norm, as well as several others including the ℓ^∞ norm. We will show that the set of piecewise regular gauges is closed under addition and pre-composition by a linear operator, which allows to cover mixed regularization (e.g. sparse+low-rank), and the so-called analysis-type priors (e.g. total variation, fused Lasso, trace Lasso, bounded polyhedral gauges). Our main result presents a unified sharp analysis of exact and robust recovery of the low-dimensional subspace model associated to the object to recover from partial measurements. This analysis is illustrated on a number of special and previously studied cases.

Keywords: Convex regularization, Inverse problems, Piecewise regular gauge, Model selection, Sparsity, Noise robustness.

1. Introduction

1.1. Regularization of Linear Inverse Problems

Inverse problems are encountered in various areas throughout science and engineering. The goal is to provably recover the structure underlying an object $x_0 \in \mathbb{R}^N$, either exactly or to a good approximation, from the partial measurements

$$y = \Phi x_0 + w, \quad (1)$$

where $y \in \mathbb{R}^Q$ is the vector of observations, $w \in \mathbb{R}^Q$ stands for the noise, and $\Phi \in \mathbb{R}^{Q \times N}$ is a linear operator which maps the N -dimensional signal domain onto the Q -dimensional observation domain. The operator Φ is in general ill-conditioned or singular, so that solving for an accurate approximation of x_0 from (1) is ill-posed.

The situation however changes if one imposes some prior knowledge on the underlying object x_0 , which makes the search for solutions to (1) feasible. This can be achieved via regularization which plays a fundamental role in bringing back ill-posed inverse problems to the land of well-posedness. We here consider solutions to the regularized optimization problem

$$x^* \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} \frac{1}{2} \|y - \Phi x\|^2 + \lambda J(x), \quad (\mathcal{P}_\lambda(y))$$

where the first term translates the fidelity of the forward model to the observations, and J is the regularization term intended to promote solutions conforming to some notion of simplicity/low-dimensional structure, that is made precise later. The regularization parameter $\lambda > 0$ is adapted to balance between the allowed fraction of noise level and regularity as dictated by the prior on x_0 . Before proceeding with the rest, it is worth mentioning that although we focus our analysis on the penalized form $(\mathcal{P}_\lambda(y))$, our results can be extended with minor adaptations to the constrained formulation, i.e. the one where the data fidelity is put as a constraint. Note also that we focus our attention for simplicity to an ℓ^2 fidelity, although our analysis carries over to more general smooth and strongly convex fidelity terms.

When there is no noise in the observations, i.e. $w = 0$ in (1), the equality-constrained minimization problem should be solved

$$x^* \in \underset{x \in \mathbb{R}^N}{\text{Argmin}} J(x) \quad \text{subject to} \quad \Phi x = y. \quad (\mathcal{P}_0(y))$$

1.2. Gauges and Model Selection

In this paper, we consider the general case where J is a *gauge* associated to a convex set C containing the origin as an *interior point*, or equivalently, that J is a gauge whose domain is full. In plain words, J is a convex, non-negative, continuous (hence bounded) and positively homogeneous function (see Section 2.1 for details).

This class of regularizers J include many well-studied ones in the literature. Among them, one can think of the ℓ^1 norm used to enforce sparse solutions [38], the discrete

total variation semi-norm [35], the $\ell^1 - \ell^2$ norm to induce block/group sparsity [44], the nuclear norm for low-rank matrices [16, 8], or bounded polyhedral gauges [41].

The behavior of this class of regularizers is dictated by the geometry of C , which is nothing in this case but the sublevel set at 1 of J , i.e., $C = \{x \in \mathbb{R}^N : J(x) \leq 1\}$, see Lemma 1. In particular, each point at which C is non-smooth encodes some low-dimensional subspace model. The regularization J then promotes solutions living in one of these subspaces. A typical example is that of the ℓ^1 -norm where the promoted vectors are sparse, living on low-dimensional subspaces aligned with the axes.

Assuming that the gauges enjoy some piecewise regularity properties, our goal in this paper is to provide a unified analysis of exact and robust recovery guarantees of the subspace model underlying the object x_0 by solving $(\mathcal{P}_\lambda(y))$ from the partial measurements in $(\mathcal{P}_\lambda(y))$. As a by-product, this entails a control on the $\|x^* - x_0\|$, where x^* is the unique minimizer of $(\mathcal{P}_\lambda(y))$.

1.3. Contributions

Our main contributions are as follows.

1.3.1. Subdifferential Decomposability of Gauges

Building upon Definition 5, which introduces the model subspace T_x at x , we provide an equivalent description of the subdifferential of a bounded gauge at x in Theorem 1. Such a description isolates and highlights a key property of a regularizing gauge, namely *decomposability*. In turn, this property allows to rewrite the first-order minimality conditions of $(\mathcal{P}_\lambda(y))$ and $(\mathcal{P}_0(y))$ in a convenient and compact way, and this lays the foundations of our subsequent contributions.

1.3.2. Uniqueness

In Theorem 2, we state a sharp sufficient condition, dubbed the Strong Null Space Property, to ensure that the solution of $(\mathcal{P}_\lambda(y))$ or $(\mathcal{P}_0(y))$ is unique. In Corollary 1, we provide a weaker sufficient condition, stated in terms of a dual vector, the existence of which certifies uniqueness. Putting together Theorem 1 and Corollary 1, Theorem 3 states the sufficient uniqueness condition in terms of a specific dual certificate built from $(\mathcal{P}_\lambda(y))$ and $(\mathcal{P}_0(y))$.

1.3.3. Piecewise Regular Gauges

In the quest for establishing robust recovery of the subspace model T_{x_0} , we first need to quantify the stability of the subdifferential of the regularizer J to local perturbations of its argument. Thus, to handle such a change of geometry, we introduce the notion of *piecewise regular gauge* (see Definition 9).

We show in particular that two important operations preserve piecewise regularity. In Proposition 8 and Proposition 10, we show that piecewise regularity of gauges is preserved under addition and pre-composition by a linear operator. Consequently, more

intricate regularizers can be built starting from simple gauges, e.g. ℓ^1 -norm, nuclear norm, etc., which are known to be piecewise regular (see the review given in Section 6).

1.3.4. Exact and Robust Subspace Recovery

This is the core contribution of the paper. Assuming the gauge is piecewise regular, we show in Theorem 5 that under a generalization of the irrepresentable condition [17], and with the proviso that the noise level is bounded and the minimal signal-to-noise ratio is high enough, there exists a whole range of the parameter λ for which problem $(\mathcal{P}_\lambda(y))$ has a unique solution x^* living in the same subspace as x_0 . In turn, this yields a control on ℓ^2 -recovery error within a factor of the noise level, i.e. $\|x^* - x_0\| = O(\|w\|)$. In the noiseless case, the irrepresentable condition implies that x_0 is exactly identified by solving $(\mathcal{P}_0(y))$.

1.4. Related Work

In [7], the authors introduced the notion of decomposable norms. In fact, we show that their regularizers are a very special subclass of ours that corresponds to strong decomposability in the sense of the Definition 8, beside symmetry since norms are symmetric gauges. Moreover, their definition involves two conditions, the second of which turns out to be an intrinsic property of gauges rather than an assumption; see the discussion after Proposition 6. Typical examples of (strongly) decomposable norms are the ℓ^1 , $\ell^1 - \ell^2$ and nuclear norms. However, strong decomposability excludes many important cases. One can think of analysis-type semi-norms since strong decomposability is not preserved under pre-composition by a linear operator, or the ℓ^∞ norm among many others. The analysis provided in [7] deals only with identifiability in the noiseless case. Their work was extended in [29] when J is the sum of decomposable norms.

Arguments based on Gaussian width were used in [9] to provide sharp estimates of the number of generic measurements required for exact and ℓ_2 -stable recovery of atomic set models from random partial information by solving a constrained form of $(\mathcal{P}_\lambda(y))$ regularized by an atomic norm. The atomic norm framework was then exploited in [31] in the particular case of the group Lasso and union of subspace models. This is however restricted to the compressed sensing scenario.

A notion of decomposability closely related to that of [7], but different, was proposed in [27]. There, the authors study ℓ_2 -stability for this class of decomposable norms with a general sufficiently smooth data fidelity. This work however only handles norms, and their stability results require however stronger assumptions than ours (typically a restricted strong convexity which becomes a type of restricted eigenvalue property for linear regression with quadratic data fidelity).

In the inverse problems literature, convergence (stability) rates have been derived in [5] with respect to the Bregman divergence for general convex regularizations J . The author in [18] established a stability result for general sublinear functions J . The stability is however measured in terms of J , and ℓ_2 -stability can only be obtained if J is coercive, which, again, excludes a large class of gauges. In [15], a ℓ_2 -stability result for decomposable norms (in the sense of [7]) precomposed by a linear operator is proved.

However, none of these works deals with exact and robust recovery of the subspace model underlying x_0 .

1.5. Paper Organization

The outline of the paper is the following. Section 2 gives essential properties of gauges and their polars, and then fully characterizes the canonical decomposition of the subdifferential of a gauge with respect to the subspace model at x . Sufficient conditions ensuring uniqueness of the minimizers to $(\mathcal{P}_\lambda(y))$ and $(\mathcal{P}_0(y))$ are provided in Section 3. In Section 4, we introduce the notion of a piecewise regular gauge and show that this property is preserved under addition and pre-composition by a linear operator. Section 5 is dedicated to our main result, namely theoretical guarantees for exact subspace recovery in presence of noise, and identifiability in the noiseless case. Section 6 exemplifies our results on several previously studied priors, and a detailed discussion on the relation with respect to relevant previous work is provided. Some conclusions and possible perspectives of this work are drawn in Section 7. The proofs of our results are collected in the appendix.

1.6. Notations and Elements from Convex Analysis

In the following, if T is a vector space, P_T denotes the orthogonal projector on T , and

$$x_T = P_T(x) \quad \text{and} \quad \Phi_T = \Phi P_T.$$

For a subset I of $\{1, \dots, N\}$, we denote by I^c its complement, $|I|$ its cardinality. $x_{(I)}$ is the subvector whose entries are those of x restricted to the indices in I , and $\Phi_{(I)}$ the submatrix whose columns are those of Φ indexed by I . For any matrix A , A^* denotes its adjoint matrix and A^+ its Moore–Penrose pseudo-inverse. We denote the right-completion of the real line by $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$.

A real-valued function $f : \mathbb{R}^N \rightarrow \overline{\mathbb{R}}$ is coercive, if $\lim_{\|x\| \rightarrow +\infty} f(x) = +\infty$. The effective domain of f is defined by $\text{dom } f = \{x \in \mathbb{R}^N : f(x) < +\infty\}$ and f is proper if $\text{dom } f \neq \emptyset$. We say that a real-valued function f is lower semi-continuous (lsc) if $\liminf_{z \rightarrow x} f(z) \geq f(x)$. A function is said sublinear if it is convex and positively homogeneous.

Let the kernel of a function be denoted $\text{Ker } f = \{x \in \mathbb{R}^N : f(x) = 0\}$. $\text{Ker } f$ is a convex cone when f is positively homogeneous.

We now provide some elements from convex analysis that are necessary throughout this paper. A comprehensive account can be found in [33, 20].

Sets For a non-empty set $C \subset \mathbb{R}^N$, we denote $\overline{\text{co}}(C)$ the closure of its convex hull. Its *affine hull* $\text{aff } C$ is the smallest affine manifold containing it, i.e.

$$\text{aff } C = \left\{ \sum_{i=1}^k \rho_i x_i : k > 0, \rho_i \in \mathbb{R}, x_i \in C, \sum_{i=1}^k \rho_i = 1 \right\}.$$

It is included in the *linear hull* $\text{span } C$ which is the smallest subspace containing C .

The interior of C is denoted $\text{int } C$. The *relative interior* $\text{ri } C$ of a convex set C is the interior of C for the topology relative to its affine hull.

Functions Let C a nonempty convex subset of \mathbb{R}^N . The *indicator function* ι_C of C is

$$\iota_C(x) = \begin{cases} 0, & \text{if } x \in C, \\ +\infty, & \text{otherwise.} \end{cases}$$

The Legendre-Fenchel *conjugate* of a proper, lsc and convex function f is

$$f^*(u) = \sup_{x \in \text{dom } f} \langle u, x \rangle - f(x),$$

where f^* is proper, lsc and convex, and $f^{**} = f$. For instance, the conjugate of the indicator function ι_C is the *support function* of C

$$\sigma_C(u) = \sup_{x \in C} \langle u, x \rangle.$$

σ_C is sublinear, is non-negative if $0 \in C$, and is finite everywhere if, and only if, C is a bounded set.

Let f and g be two functions proper closed convex functions from \mathbb{R}^N to $\overline{\mathbb{R}}$. Their infimal convolution is the function

$$(f \check{\vee} g)(x) = \inf_{x_1 + x_2 = x} f(x_1) + g(x_2) = \inf_{z \in \mathbb{R}^N} f(z) + g(x - z).$$

The *subdifferential* $\partial f(x)$ of a convex function f at x is the set

$$\partial f(x) = \{u \in \mathbb{R}^N : f(x') \geq f(x) + \langle u, x' - x \rangle, \quad \forall x' \in \text{dom } f\}.$$

An element of $\partial f(x)$ is a subgradient. If the convex function f is Gâteaux-differentiable at x , then its only subgradient is its gradient, i.e. $\partial f(x) = \{\nabla f(x)\}$.

The *directional derivative* $f'(x, \delta)$ of a lsc function f at the point $x \in \text{dom } f$ in the direction $\delta \in \mathbb{R}^N$ is

$$f'(x, \delta) = \lim_{t \downarrow 0} \frac{f(x + t\delta) - f(x)}{t}.$$

When f is convex, then the function $\delta \mapsto f'(x, \delta)$ exists and is sublinear. The subdifferential $\partial f(x)$ is a non-empty compact convex set of \mathbb{R}^N whose support function is $f'(x, \cdot)$, i.e.

$$f'(x, \delta) = \sigma_{\partial f(x)}(\delta) = \sup_{\eta \in \partial f(x)} \langle \eta, \delta \rangle.$$

We also recall the fundamental first-order minimality condition of a convex function: x^* is the global minimizer of a convex function f if, and only if, $0 \in \partial f(x)$.

2. Geometry of Gauge Regularization

2.1. Gauges and their Polars

Definitions and main properties We start by collecting some important properties of gauges and their polars. A comprehensive account on them can be found in [33].

We begin with the definition of a gauge.

Definition 1 (Gauge). *Let $C \subseteq \mathbb{R}^N$ be a non-empty closed convex set containing the origin. The gauge of C is the function γ_C defined on \mathbb{R}^N by*

$$\gamma_C(x) = \inf \{ \lambda > 0 : x \in \lambda C \}.$$

As usual, $\gamma_C(x) = +\infty$ if the infimum is not attained.

Lemma 1.

- (i) γ_C is a non-negative, lsc and sublinear function.
- (ii) C is the unique closed convex set containing the origin such that

$$C = \{x \in \mathbb{R}^N : \gamma_C(x) \leq 1\}.$$

- (iii) γ_C is bounded if, and only if, $0 \in \text{int } C$, in which case γ_C is continuous.
- (iv) $\text{Ker } \gamma_C = \{0\}$, or equivalently γ_C is coercive if, and only if, C is compact.
- (v) γ_C is bounded and coercive on $\text{dom } \gamma_C = \text{span } C$ if, and only if, C is compact and $0 \in \text{ri } C$. In particular, γ_C is bounded and coercive if, and only if, C is compact and $0 \in \text{int } C$.

Lemma 1(ii) is fundamental result of convex analysis that states that there is a one-to-one correspondence between gauge functions and closed convex sets containing the origin. This allows to identify sets from their gauges, and vice versa. Recall that in this paper, we consider a regularizing gauge J associated to a convex set C containing the origin as an interior point, or equivalently by Lemma 1(i)-(iii), that J is a non-negative, continuous and sublinear function of full domain.

γ_C is a norm, having C as its unit ball, if and only if C is bounded with nonempty interior and symmetric. When C is only symmetric with nonempty interior, then γ_C becomes a semi-norm.

Let us now turn to the polar of a convex set and a gauge.

Definition 2 (Polar set). *Let C be a non-empty convex set. The set C° given by*

$$C^\circ = \{v \in \mathbb{R}^N : \langle v, x \rangle \leq 1 \text{ for all } x \in C\}$$

is called the polar of C .

C° is a closed convex set containing the origin. When the set C is also closed and contains the origin, then it coincides with its bipolar, i.e. $C^{\circ\circ} = C$.

We are now in position to define the polar gauge.

Definition 3 (Polar Gauge). *The polar of a gauge γ_C is the function γ_C° defined by*

$$\gamma_C^\circ(u) = \inf \{ \mu \geq 0 : \langle x, u \rangle \leq \mu \gamma_C(x), \forall x \} .$$

Observe that gauges polar to each other have the property

$$\langle x, u \rangle \leq \gamma_C(x) \gamma_C^\circ(u) \quad \forall (x, u) \in \text{dom } \gamma_C \times \text{dom } \gamma_C^\circ ,$$

just as dual norms satisfy a duality inequality. In fact, polar pairs of gauges correspond to the best inequalities of this type.

Lemma 2. *Let $C \subseteq \mathbb{R}^N$ be a closed convex set containing 0. Then,*

(i) γ_C° is a gauge function and $\gamma_{C^\circ} = \gamma_C$.

(ii) $\gamma_C^\circ = \gamma_{C^\circ}$, or equivalently

$$C^\circ = \{x \in \mathbb{R}^N : \gamma_C^\circ(x) \leq 1\} = \{x \in \mathbb{R}^N : \gamma_{C^\circ}(x) \leq 1\} .$$

(iii) *The gauge of C and the support function of C are mutually polar, i.e.*

$$\gamma_C = \sigma_{C^\circ} \quad \text{and} \quad \gamma_{C^\circ} = \sigma_C .$$

Gauge and polar calculus We here derive the expression of the gauge function of the Minkowski sum of two sets, as well as that of the image of a set by a linear operator. These results play an important role in Section 4.

Lemma 3. *Let C_1 and C_2 be nonempty closed convex sets containing the origin. Then*

$$\gamma_{C_1+C_2}(x) = \sup_{\rho \in [0,1]} \rho \gamma_{C_1} \overset{+}{\vee} (1-\rho) \gamma_{C_2}(x) .$$

If x is such that $\gamma_{C_1}(x_1) + \gamma_{C_2}(x_2)$ is continuous and bounded on $\{(x_1, x_2) : x_1 + x_2 = x\}$, then

$$\gamma_{C_1+C_2}(x) = \inf_{z \in \mathbb{R}^N} \max(\gamma_{C_1}(z), \gamma_{C_2}(x-z)) .$$

Lemma 4. *Let C be a closed convex set such that $0 \in \text{ri } C$, and D a linear operator. Then, for every $x \in \text{Im}(D)$*

$$\gamma_{D(C)}(x) = \inf_{z \in \text{Ker}(D)} \gamma_C(D^+x + z) .$$

Using Lemma 1(v), one can observe that the infimum is bounded if $(D^+x + \text{Ker}(D)) \cap \text{span } C \neq \emptyset$.

Operator bound

Definition 4. Let J_1 and J_2 be two gauges defined on two vector spaces V_1 and V_2 , and $A : V_1 \rightarrow V_2$ a linear map. The operator bound $\mathcal{M}_{J_1, J_2}(A)$ of A between J_1 and J_2 is given by

$$\mathcal{M}_{J_1, J_2}(A) = \sup_{J_1(x) \leq 1} J_2(Ax).$$

Note that $\mathcal{M}_{J_1, J_2}(A) < +\infty$ if, and only if $A \text{Ker}(J_1) \subseteq \text{Ker}(J_2)$. In particular, if J_1 is coercive (i.e. $\text{Ker } J_1 = \{0\}$ from Lemma 1(iv)), then $\mathcal{M}_{J_1, J_2}(A)$ is finite. As a convention, $\mathcal{M}_{J_1, \|\cdot\|_p}(A)$ is denoted as $\mathcal{M}_{J_1, \ell^p}(A)$. An easy consequence of this definition is the fact that for every $x \in V_1$,

$$J_2(Ax) \leq \mathcal{M}_{J_1, J_2}(A) J_1(x).$$

We end this section by pointing out that many of the results proved in this paper can be extended to f -homogenous closed convex functions J , i.e. $J(\lambda x) = f(\lambda)J(x)$, $\forall x$ and $\lambda > 0$ for a positive continuous increasing convex function f on \mathbb{R}_+ . These are gauge-like functions that can be built from gauges as $J = g \circ \gamma_C$, where g is a non-constant non-decreasing and lsc convex function on \mathbb{R}_+ such that $g(\lambda t) = f(\lambda)g(t)$. This construction can be proved from [33, Theorem 15.3]. An important class is that of positively homogenous functions of degree p , in which case $g(t) = t^p/p$ and $f(\lambda) = \lambda^p$ for $1 \leq p < \infty$. One can show that the main robust recovery result proved in this paper can be extended to f -homogenous closed convex functions, at the price of modifying the scaling between the regularization parameter λ and the noise level.

2.2. Model Subspace Associated to a Gauge

Let J our regularizer, i.e. a bounded gauge. We now introduce the model subspace at a point x .

Definition 5 (Model Subspace). For any vector $x \in \mathbb{R}^N$, we denote by \bar{S}_x the affine hull of the subdifferential of J at x

$$\bar{S}_x = \text{aff } \partial J(x),$$

and e_x the orthogonal projection of 0 onto \bar{S}_x

$$e_x = \underset{e \in \bar{S}_x}{\text{argmin}} \|e\|.$$

We denote

$$S_x = \bar{S}_x - e_x = \text{span}(\partial J(x) - e_x) \quad \text{and} \quad T_x = S_x^\perp.$$

T_x is coined the model subspace of x associated to J .

When J is Gâteaux-differentiable at x , i.e. $\partial J(x) = \{\nabla J(x)\}$, $e_x = \nabla J(x)$ and $T_x = \mathbb{R}^N$. On the contrary, when J is not smooth at x , the dimension of T_x is smaller dimension, and the regularizing gauge J essentially promotes elements living on this model subspace.

We start by summarizing some key properties of the objects e_x and T_x .

Proposition 1. *For any $x \in \mathbb{R}^N$, one has*

- (i) *For every $u \in \bar{S}_x$, $J(x) = \langle u, x \rangle$.*
- (ii) *$e_x \in T_x \cap \bar{S}_x$.*
- (iii) *$\bar{S}_x = \{\eta \in \mathbb{R}^N : \eta_{T_x} = e_x\}$.*
- (iv) *$x \in T_x$.*

In general $e_x \notin \partial J(x)$, which is the situation displayed on Figure 1. It is worth noting that the fact that $e_x \in T_x \cap \bar{S}_x$ holds for any proper convex function, not only gauges, whereas $x \in T_x$ does not hold in general.

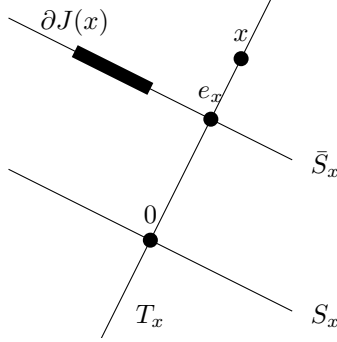


Figure 1: Illustration of the geometrical elements (S_x, T_x, e_x) .

From this section until Section 4, we use the ℓ^1 - ℓ^2 and the ℓ^∞ norms as illustrative examples. A more comprehensive treatment is provided in Section 6 completely dedicated to examples.

Examples

Example 1 (ℓ^1 - ℓ^2 norm). We consider a uniform disjoint partition \mathcal{B} of $\{1, \dots, N\}$,

$$\{1, \dots, N\} = \bigcup_{b \in \mathcal{B}} b, \quad b \cap b' = \emptyset, \quad \forall b \neq b'.$$

The $\ell^1 - \ell^2$ norm of x is

$$J(x) = \|x\|_{\mathcal{B}} = \sum_{b \in \mathcal{B}} \|x_b\|.$$

The subdifferential of J at $x \in \mathbb{R}^N$ is

$$\partial J(x) = \left\{ \eta \in \mathbb{R}^N : \forall b \in I, \eta_b = \frac{x_b}{\|x_b\|} \text{ and } \forall g \notin I, \|\eta_g\| \leq 1 \right\},$$

where $I = \{b \in \mathcal{B} : x_b \neq 0\}$. Thus, the affine hull of $\partial J(x)$ reads

$$\bar{S}_x = \left\{ \eta \in \mathbb{R}^N : \forall b \in I, \eta_b = \frac{x_b}{\|x_b\|} \right\}.$$

Hence the projection of 0 onto \bar{S}_x is

$$e_x = (\mathcal{N}(x_b))_{b \in \mathcal{B}}$$

where $\mathcal{N}(a) = a/\|a\|$ if $a \neq 0$, and $\mathcal{N}(0) = 0$ and

$$S_x = \bar{S}_x - e_x = \left\{ \eta \in \mathbb{R}^N : \forall b \in I, \eta_b = 0 \right\},$$

and

$$T_x = S_x^\perp = \left\{ \eta \in \mathbb{R}^N : \forall b \notin I, \eta_b = 0 \right\}.$$

Example 2 (ℓ^∞ norm). The ℓ^∞ norm is $J(x) = \|x\|_\infty = \max_{1 \leq i \leq N} |x_i|$.

For $x = 0$, $\partial J(x)$ is the unit ℓ^1 ball, hence $\bar{S}_x = S_x = \mathbb{R}^N$, $T_x = \{0\}$ and $e_x = 0$.

For $x \neq 0$, we have

$$\partial J(x) = \left\{ \eta : \forall i \in I^c, \eta_i = 0, \langle \eta, s \rangle = 1, \eta_i s_i > 0 \forall i \in I \right\}.$$

where $I = \{i \in \{1, \dots, N\} : |x_i| = \|x\|_\infty\}$ and $s = \text{sign}(x)$, with $\text{sign}(0) = 0$. It is clear that \bar{S}_x is the affine hull of an $|I|$ -dimensional face of the unit ℓ^1 ball exposed by the sign subvector $s_{(I)}$. Thus e_x is the barycenter of that face, i.e.

$$e_x = s/|I| \text{ and } S_x = \left\{ \eta : \eta_{(I^c)} = 0 \text{ and } \langle \eta_{(I)}, s_{(I)} \rangle = 0 \right\}.$$

In turn

$$T_x = S_x^\perp = \left\{ \alpha : \alpha_{(I)} = \rho s_{(I)} \text{ for } \rho \in \mathbb{R} \right\}.$$

2.3. Decomposability Property

The subdifferential of a gauge γ_C at a point x is completely characterized by the face of its polar set C° exposed by x . Put formally, we have [20]

$$\partial \gamma_C(x) = \mathbf{F}_{C^\circ}(x) = \left\{ \eta \in \mathbb{R}^N : \eta \in C^\circ \text{ and } \langle \eta, x \rangle = \gamma_C(x) \right\},$$

where $\mathbf{F}_{C^\circ}(x)$ is the face of C° exposed by x . The latter is the intersection of C° and the supporting hyperplane $\{\eta \in \mathbb{R}^N : \langle \eta, x \rangle = \gamma_C(x)\}$. The special case of $x = 0$ has a much simpler structure; it is the polar set C° from Lemma 2(ii)-(iii), i.e.

$$\partial \gamma_C(x) = \left\{ \eta \in \mathbb{R}^N : \gamma_{C^\circ}(\eta) \leq 1 \right\} = C^\circ.$$

The following proposition gives an equivalent convenient description of the subdifferential of the regularizer $J = \gamma_C$ at x in terms of a particular supporting hyperplane to C° : the affine hull \bar{S}_x .

Proposition 2. *Let $J = \gamma_C$ be a bounded gauge. Then for $x \in \mathbb{R}^N$, one has*

$$\partial J(x) = \bar{S}_x \cap C^\circ.$$

The antipromoting gauge and its polar Before providing an equivalent description of the subdifferential of J at x in terms of the geometrical objects e_x , T_x and S_x , we introduce another gauge that plays a prominent role in this description.

Definition 6 (Antipromoting Gauge). *Let J be a bounded gauge, and denote C the associated closed convex set (hence containing the origin in its interior). Let $x \in \mathbb{R}^N \setminus \{0\}$ and $f_x \in \text{ri } \partial J(x)$. The antipromoting gauge associated to f_x is the gauge $J_{f_x}^\circ = \gamma_{\partial J(x) - f_x}$.*

Since $\partial J(x) - f_x$ is a closed (in fact compact) convex set containing the origin, it is uniquely characterized by the antipromoting gauge $J_{f_x}^\circ$ (see Lemma 1(i)).

The following proposition states the main properties of the gauge $J_{f_x}^\circ$.

Proposition 3. *The antipromoting gauge $J_{f_x}^\circ$ is such that $\text{dom } J_{f_x}^\circ = S_x$, and is coercive on S_x . Moreover,*

$$J_{f_x}^\circ(\eta) = \inf_{\tau \geq 0} \max(J^\circ(\tau f_x + \eta), \tau) + \iota_{S_x}(\eta).$$

The second claim gives a formula which links $J_{f_x}^\circ$ to the polar gauge J° . But they are not equal in general unless some additional assumptions are imposed on J , as we will see shortly in Section 2.4.

We now turn to the gauge polar to the anti-promoting gauge $(J_{f_x}^\circ)^\circ = J_{f_x}$, where the last equality is a consequence of Lemma 2(i). J_{f_x} comes into play in several results in the rest of the paper. The following proposition summarizes its most important properties.

Proposition 4. *The gauge J_{f_x} is such that*

- (i) *Its has a full domain.*
- (ii) $J_{f_x}(d) = J_{f_x}(d_S) = \sup_{J_{f_x}^\circ(\eta_{S_x}) \leq 1} \langle \eta_{S_x}, d \rangle$.
- (iii) $\text{Ker } J_{f_x} = T_x$ and J_{f_x} is coercive on S_x .
- (iv) $J_{f_x}(d) = J(d_{S_x}) - \langle f_{S_x}, d_{S_x} \rangle$

Let's derive the antipromoting gauge on the illustrative example of the ℓ^∞ norm. The case of $\ell^1 - \ell^2$ norm is detailed in Section 2.4.

Example 3 (ℓ^∞ norm). Recall from Section 2.2 that for $J = \|\cdot\|_\infty$, $f_x = e_x = s/|I|$, with $s = \text{sign}(x)$. Let $\mathcal{K}_x = \partial J(x) - e_x$. It can be straightforwardly shown that in this case,

$$\mathcal{K}_x = \{v : \forall (i, j) \in I \times I^c, v_j = 0, \langle v_{(I)}, s_{(I)} \rangle = 0, -|I|v_i s_i \leq 1\}$$

This is rewritten as

$$\mathcal{K}_x = S_x \cap \underbrace{\{v : \forall i \in I, -|I|v_i s_i \leq 1\}}_{=\mathcal{K}'_x}.$$

Thus the antipromoting gauge reads

$$J_{f_x}^\circ(\eta) = \gamma_{\mathcal{K}_x}(\eta) = \max(\gamma_{S_x}(\eta), \gamma_{\mathcal{K}'_x}(\eta)).$$

We have $\gamma_{S_x}(\eta) = \iota_{S_x}(\eta)$ and $\gamma_{\mathcal{K}'_x}(\eta) = \max_{i \in I} (-|I|s_i \eta_i)_+$, where $(\cdot)_+$ is the positive part, hence we obtain

$$J_{f_x}^\circ(\eta) = \begin{cases} \max_{i \in I} (-|I|s_i \eta_i)_+ & \text{if } \eta \in S_x \\ +\infty & \text{otherwise.} \end{cases}$$

Therefore the subdifferential of $\|\cdot\|_\infty$ at x takes the form

$$\partial J(x) = \left\{ \eta \in \mathbb{R}^N : \eta_{T_x} = e_x = \frac{s}{|I|} \quad \text{and} \quad \max_{i \in I} (-|I|s_i \eta_i)_+ \leq 1 \right\}.$$

Decomposability of the subdifferential Piecing together the above ingredients yields a fundamental pointwise decomposition of the subdifferential of the regularizer gauge J . This decomposability property is at the heart of our results in the rest of the paper.

Theorem 1 (Decomposability). *Let $J = \gamma_C$ be a bounded gauge. Let $x \in \mathbb{R}^N \setminus \{0\}$ and $f_x \in \text{ri } \partial J(x)$. Then the subdifferential of J at x reads*

$$\partial J(x) = \left\{ \eta \in \mathbb{R}^N : \eta_{T_x} = e_x \quad \text{and} \quad J_{f_x}^\circ(\text{P}_{S_x}(\eta - f_x)) \leq 1 \right\}.$$

First-order minimality condition Capitalizing on Theorem 1, we are now able to deduce a convenient necessary and sufficient first-order (global) minimality condition of $(\mathcal{P}_\lambda(y))$ and $(\mathcal{P}_0(y))$.

Proposition 5. *Let $x \in \mathbb{R}^N$, and denote for short $T = T_x$ and $S = S_x$. The two following propositions hold.*

(i) *The vector x is a global minimizer of $(\mathcal{P}_\lambda(y))$ if, and only if,*

$$\Phi_T^*(y - \Phi x) = \lambda e_x \quad \text{and} \quad J_{f_x}^\circ(\lambda^{-1} \Phi_S^*(y - \Phi x) - \text{P}_S(f_x)) \leq 1.$$

(ii) *The vector x is a global minimizer of $(\mathcal{P}_0(y))$ if, and only if, there exists a dual vector $\alpha \in \mathbb{R}^Q$ such that*

$$\Phi_T^* \alpha = e_x \quad \text{and} \quad J_{f_x}^\circ(\Phi_S^* \alpha - \text{P}_S(f_x)) \leq 1.$$

2.4. Strong Gauge

In this section, we study a particular subclass of gauges that we dub strong gauges. We start with some definitions.

Definition 7. A bounded regularizing gauge J is separable with respect to $T = S^\perp$ if

$$\forall (x, x') \in T \times S, \quad J(x + x') = J(x) + J(x').$$

Separability of J is equivalent to the following property on the polar J° .

Lemma 5. Let J be a bounded gauge. Then, J is separable w.r.t. to $T = S^\perp$ if, and only if its polar J° satisfies

$$J^\circ(x + x') = \max(J^\circ(x), J^\circ(x')), \quad \forall (x, x') \in T \times S.$$

The decomposability of $\partial J(x)$ as described in Theorem 1 depends on the particular choice of the map $x \mapsto f_x \in \text{ri } \partial J(x)$. An interesting situation is encountered when $e_x \in \text{ri } \partial J(x)$, in which case, one can just choose $f_x = e_x$, hence implying that $f_{S_x} = 0$. Strong gauges are precisely a class of gauges for which this situation occurs.

In the sequel, for a given model subspace T , we denote \tilde{T} the set of vectors sharing the same T ,

$$\tilde{T} = \{x \in \mathbb{R}^N : T_x = T\}.$$

Using positive homogeneity, it is easy to show that $T_{\rho x} = T_x$ and $e_{\rho x} = e_x \forall \rho > 0$. Thus \tilde{T} is a non-empty cone which is contained in T by Proposition 1(iv).

Definition 8 (Strong Gauge). A strong gauge on T is a bounded gauge J such that

1. For every $x \in \tilde{T}$, $e_x \in \text{ri } \partial J(x)$.
2. J is separable with respect to T and $S = T^\perp$.

The following result shows that the decomposability property of Theorem 1 has a simpler form when J is a strong gauge.

Proposition 6. Let J be a strong gauge on T_x . Then, the subdifferential of J at x reads

$$\partial J(x) = \{\eta \in \mathbb{R}^N : \eta_{T_x} = e_x \quad \text{and} \quad J^\circ(\eta_{S_x}) \leq 1\}.$$

When J is in addition a norm, this coincides with the decomposability definition of [7]. Note however that the last part of assertion (ii) in Proposition 4 is an intrinsic property of gauges, while it is stated as an assumption in their definition.

Example 4 (ℓ^1 - ℓ^2 norm). Recall the notations of this example in Section 2.2. Since $e_x = (\mathcal{N}(x_b))_{b \in \mathcal{B}} \in \text{ri } \partial J(x)$, and the ℓ^1 - ℓ^2 norm is separable, it is a strong norm according to Definition 8. Thus, its subdifferential at x reads

$$\partial J(x) = \left\{ \eta \in \mathbb{R}^N : \eta_{T_x} = e_x = (\mathcal{N}(x_b))_{b \in \mathcal{B}} \quad \text{and} \quad \max_{b \notin T} \|\eta_b\| \leq 1 \right\}.$$

However, except for $N = 2$, ℓ^∞ is not a strong gauge.

3. Uniqueness

This section derives sufficient conditions under which the solution of problems $(\mathcal{P}_\lambda(y))$ and $(\mathcal{P}_0(y))$ is unique.

We start with the key observation that although $(\mathcal{P}_\lambda(y))$ does not necessarily have a unique minimizer in general, all solutions share the same image under Φ .

Lemma 6. *Let x, x' be two solutions of $(\mathcal{P}_\lambda(y))$. Then,*

$$\Phi x = \Phi x'.$$

Consequently, the set of the minimizers of $(\mathcal{P}_\lambda(y))$ is a closed convex subset of the affine space $x + \text{Ker}(\Phi)$, where x is any minimizer of $(\mathcal{P}_\lambda(y))$. This is also obviously the case for $(\mathcal{P}_0(y))$ since all feasible solutions belong to the affine space $x_0 + \text{Ker} \Phi$.

3.1. The Strong Null Space Property

The following theorem gives a sufficient condition to ensure uniqueness of the solution to $(\mathcal{P}_\lambda(y))$ and $(\mathcal{P}_0(y))$, that we coin *Strong Null Space Property*. This condition is a generalization of the Null Space Property introduced in [12] and popular in ℓ^1 regularization.

Theorem 2. *Let x be a solution of $(\mathcal{P}_\lambda(y))$ (resp. $(\mathcal{P}_0(y))$). Denote $T = S^\perp = T_x$ the associated model subspace. If the Strong Null Space Property holds*

$$\forall \delta \in \text{Ker}(\Phi) \setminus \{0\}, \quad \langle e_x, \delta_T \rangle + \langle P_S(f_x), \delta_S \rangle < J_{f_x}(-\delta_S), \quad (\text{NSP}^S)$$

then the vector x is the unique minimizer of $(\mathcal{P}_\lambda(y))$ (resp. $(\mathcal{P}_0(y))$).

This result reduces to the one proved in [15] when J is decomposable norm pre-composed by a linear operator, where decomposability is intended in the sense of [7]. This is covered by our result when specializing it to a strong gauge J . In such a case, (NSP^S) reads

$$\forall \delta \in \text{Ker}(\Phi) \setminus \{0\}, \quad \langle e_x, \delta_{T_x} \rangle < J(-\delta_{S_x}).$$

3.2. Dual Certificates

In this section we derive from (NSP^S) a weaker sufficient condition, stated in terms of a dual vector, the existence of which certifies uniqueness.

For some model subspace T , the restricted injectivity of Φ on T plays a central role in the sequel. This is achieved by imposing that

$$\text{Ker}(\Phi) \cap T = \{0\}. \quad (\mathcal{C}_T)$$

To understand the importance of (\mathcal{C}_T) , consider the noiseless case where we want to recover a vector x_0 from $y = \Phi x_0$, whose model subspace is T . Assume that the latter is known. From Proposition 1(iv), $x_0 \in T \cap \{x : y = \Phi x\}$. For x_0 to be uniquely

recovered from y , (\mathcal{C}_T) must be verified. Otherwise, if (\mathcal{C}_T) does not hold, then any $x_0 + \delta$, with $\delta \in \text{Ker } \Phi \cap T \setminus \{0\}$, is also a candidate solution. Thus, such objects cannot be uniquely recovered.

We can derive from Theorem 2 the following corollary.

Corollary 1. *Let x be a solution of $(\mathcal{P}_\lambda(y))$ (resp. $(\mathcal{P}_0(y))$). Assume that there exists a dual vector α such that $\eta = \Phi^* \alpha \in \text{ri}(\partial J(x))$, and (\mathcal{C}_T) holds where $T = T_x$. Then x is the unique solution of $(\mathcal{P}_\lambda(y))$ (resp. $(\mathcal{P}_0(y))$).*

Piecing together Theorem 1 and Corollary 1, one can build a particular dual certificate for $(\mathcal{P}_\lambda(y))$, and then state a sufficient uniqueness explicitly in terms of the decomposable structure of the subdifferential of the regularizing gauge J .

Theorem 3. *Let $x \in \mathbb{R}^N$, and suppose that $f_x \in \text{ri} \partial J(x)$. Assume furthermore that (\mathcal{C}_T) holds for $T = T_x$ and let $S = T^\perp$.*

(i) *If*

$$\Phi_T^*(y - \Phi x) = \lambda e_x, \quad (2)$$

$$J_{f_x}^\circ (\lambda^{-1} \Phi_S^*(y - \Phi x) - P_S(f_x)) < 1. \quad (3)$$

then x is the unique solution of $(\mathcal{P}_\lambda(y))$.

(ii) *If there exists a dual certificate α such that*

$$\Phi_T^* \alpha = e_x \quad \text{and} \quad J_{f_x}^\circ (\Phi_S^* \alpha - P_S(f_x)) < 1.$$

then x is the unique solution of $(\mathcal{P}_0(y))$.

4. Piecewise Regular Gauges

Until now, except of being bounded (i.e. full domain), no other assumption was imposed on the regularizing gauge J . But, toward the goal of studying robust recovery by solving $(\mathcal{P}_\lambda(y))$, more will be needed. This is the main reason underlying the introduction of a subclass of bounded gauges J for which the mappings $x \mapsto e_x$, $x \mapsto P_{S_x}(f_x)$ and $x \mapsto J_{f_x}^\circ$ exhibit local regularity.

4.1. Piecewise Regular Gauges

Definition 9 (Piecewise Regular Gauge). *A bounded gauge J is said to be piecewise regular (PRG for short) at $x \in \mathbb{R}^N$, with respect to the mapping $x \mapsto f_x \in \text{ri} \partial J(x)$ and Γ a coercive gauge bounded on $T_x = T = S^\perp$, if there exist three non-negative reals ν_x, μ_x, τ_x and a real ξ_x such that for every $x' \in T$ with $\Gamma(x - x') \leq \nu_x$, one has*

$T_x = T_{x'}$ (i.e. $x' \in \tilde{T}$) and

$$\Gamma(e_x - e_{x'}) \leq \mu_x \Gamma(x - x'). \quad (4)$$

$$J_{f_x}^\circ(\mathbb{P}_S(f_x - f_{x'})) \leq \tau_x \Gamma(x - x'). \quad (5)$$

$$\sup_{\substack{u \in S \\ u \neq 0}} \frac{J_{f_{x'}}^\circ(u) - J_{f_x}^\circ(u)}{J_{f_x}^\circ(u)} \leq \xi_x \Gamma(x - x'). \quad (6)$$

The gauge J is strongly piecewise regular at $x \in \mathbb{R}^N$ if it is piecewise regular at x and strong (Definition 8) with respect to T .

In plain words, (4), (5) and (6) amount to imposing that J is sufficiently regular (Lipschitzian-like) on \tilde{T} . Recall from Section 2.4 that \tilde{T} is a non-empty cone and $e_{\rho x} = e_x$ for all $\rho > 0$. Thus we have $\mu_{\rho x} = \rho \mu_x$. Moreover, one can always impose $\mu_x \nu_x \leq C$ for some global constant C (but of course ν_x can be smaller and even zero).

Many well-studied gauges are strongly piecewise regular. One can for instance think of the ℓ^1 -norm, the $\ell^1 - \ell^2$ -norm and the nuclear norm. The analysis- ℓ^1 regularization or ℓ^∞ -norm are however not strong PRG. Piecewise regularity of these examples is investigated in details in Section 6.

4.2. Operations Preserving Piecewise Regularity

Piecewise regularity is preserved under addition and pre-composition by a linear operator. It is worth observing that strong piecewise regularity is however not preserved under the same operations. More precisely, the composition of a strong gauge with a linear operator is not strong anymore, nor is the gauge of the sum of two strong gauges.

4.2.1. Addition

The following proposition determines the model subspace and the antipromoting gauge of the sum of two gauges

$$H = J + G$$

in terms of those associated to J and G .

Proposition 7. *Let J and G be two bounded gauges. Denote T^J and e_J (resp. T^G and e_G) the model subspace and vector at a point x corresponding to J (resp. G). Then the subdifferential of H has the decomposability property with*

$$(i) \quad T^H = T^J \cap T^G, \text{ or equivalently } S^H = (T^H)^\perp = \text{span}(S^J \cup S^G).$$

$$(ii) \quad e_H = \mathbb{P}_{T^H}(e_J + e_G).$$

(iii) *Moreover, let $J_{f_x^J}^\circ$ and $G_{f_x^G}^\circ$ denote the antipromoting gauges for the pairs $(J, f_x^J \in \text{ri } \partial J(x))$ and $(G, f_x^G \in \text{ri } \partial G(x))$, correspondingly. Then, for the particular choice of*

$$f_x^H = f_x^J + f_x^G$$

we have $f_x^H \in \text{ri} \partial H(x)$, and for a given $\eta \in S^H$, the antipromoting gauge of H reads

$$H_{f_x^H}^\circ(\eta) = \inf_{\eta_1 + \eta_2 = \eta} \max(J_{f_x^J}^\circ(\eta_1), G_{f_x^G}^\circ(\eta_2)) .$$

We now establish piecewise regularity of the sum with the decomposability property dictated by Proposition 7.

Proposition 8. *If J and G are piecewise regular gauges at x with the corresponding parameters $(\Gamma^J, \nu_x^J, \mu_x^J, \tau_x^J, \xi_x^J)$ and $(\Gamma^G, \nu_x^G, \mu_x^G, \tau_x^G, \xi_x^G)$, then $H = J + G$ is also piecewise regular at x , for the choice $f_x^H = f_x^J + f_x^G$ and $\Gamma^H = \max(\Gamma^J, \Gamma^G)$, with the parameters*

$$\begin{aligned} \nu_x^H &= \min(\nu_x^J, \nu_x^G) \\ \mu_x^H &= \mu_x^J \mathcal{M}_{\Gamma^J, \Gamma^H}(\mathbb{P}_{T^H}) + \mu_x^G \mathcal{M}_{\Gamma^G, \Gamma^H}(\mathbb{P}_{T^H}) \\ \tau_x^H &= \tau_x^J + \tau_x^G + \mu_x^J \mathcal{M}_{\Gamma^J, H_{f_x^H}^\circ}(\mathbb{P}_{S^H \cap T^J}) + \mu_x^G \mathcal{M}_{\Gamma^G, H_{f_x^H}^\circ}(\mathbb{P}_{S^H \cap T^G}) \\ \xi_x^H &= \max(\xi_x^J, \xi_x^G). \end{aligned}$$

4.2.2. Pre-composition by a Linear Operator

Gauges of the form $J_0 \circ D^*$, where J_0 is a bounded regularizing gauge, correspond to the so-called analysis-type regularizers. The most popular example in this class is the total variation where J_0 is the ℓ^1 or the $\ell^1 - \ell^2$ norm, and $D^* = \nabla$ is a finite difference discretization of the gradient.

In the following, we denote $T = T_x = S^\perp$ and $e = e_x$ the subspace and vector in the decomposition of the subdifferential of J at a given $x \in \mathbb{R}^N$. Analogously, $T_0 = S_0^\perp$ and e_0 are those of the gauge J_0 at D^*x . The following proposition details the decomposability structure of analysis-type gauges.

Proposition 9. *Let J_0 be a bounded gauge. With the above notation, the subdifferential of the analysis gauge $J = J_0 \circ D^*$, which is also bounded, has the decomposability property with*

- (i) $T = \text{Ker}(D_{S_0}^*)$, or equivalently $S = \text{Im}(D_{S_0})$.
- (ii) $e = P_T D e_0$.
- (iii) Moreover, let J_{0, f_0, D^*x}° denote the antipromoting gauge for the pair $(J_0, f_{0, D^*x} \in \text{ri} \partial J_0(x))$. Then, for the particular choice of

$$f_x = D f_{0, D^*x}$$

we have $f_x \in \text{ri} \partial J(x)$, $\text{dom} J_{f_x}^\circ = S$ and for every $\eta \in S$

$$J_{f_x}^\circ(\eta) = \inf_{z \in \text{Ker}(D_{S_0})} J_{0, f_0, D^*x}^\circ(D_{S_0}^+ \eta + z) .$$

The infimum can be equivalently taken over $\text{Ker}(D) \cap S_0$.

Capitalizing on these properties, we now establish that an analysis gauge $J = J_0 \circ D^*$, where J_0 is a PRG, is also piecewise regular.

Proposition 10. *If J_0 is a bounded piecewise regular gauge at $u = D^*x$ with the parameters $(\Gamma_0, \nu_{0,u}, \mu_{0,u}, \tau_{0,u}, \xi_{0,u})$, then $J = J_0 \circ D^*$ is piecewise regular at x , with the choice $f_x = Df_{0,u}$ and Γ any bounded coercive gauge on T , with the parameters*

$$\begin{aligned}\nu_x &= \frac{1}{\mathcal{M}_{\Gamma, \Gamma_0}(D^*)} \nu_{0,u} \\ \mu_x &= \mu_{0,u} \mathcal{M}_{\Gamma, \Gamma_0}(P_T D) \mathcal{M}_{\Gamma, \Gamma_0}(D^*) \\ \tau_x &= \left(\tau_{0,u} \mathcal{M}_{J_{0,f_{0,u}}^\circ, J_{0,f_{0,u}}^\circ} (D_{S_0}^+ P_S D) + \mu_{0,u} \mathcal{M}_{\Gamma_0, J_{0,f_{0,u}}^\circ} (D_{S_0}^+ P_S D) \right) \mathcal{M}_{\Gamma, \Gamma_0}(D^*) \\ \xi_x &= \xi_{0,u} \mathcal{M}_{\Gamma, \Gamma_0}(D^*) .\end{aligned}$$

5. Exact Model Selection and Identifiability

In this section, we state our main recovery guarantee, which asserts that under appropriate conditions, $(\mathcal{P}_\lambda(y))$ with a piecewise regular gauge J has a unique solution x^* whose model subspace $T_{x^*} = T_{x_0}$, even in presence of noise. Put differently, regularization by J is able to select the correct model subspace underlying x_0 .

Beside condition (\mathcal{C}_{T_x}) stated above, the following Identifiability Criterion will play a pivotal role.

Definition 10. *For $x \in \mathbb{R}^N$ such that (\mathcal{C}_{T_x}) holds, we define the Identifiability Criterion at x as*

$$\mathbf{IC}(x) = J_x^\circ (\Phi_{S_x}^* \Phi_{T_x}^{+,*} e_x - P_{S_x} f_x).$$

Note that if J is a strong gauge on T , then it becomes $\mathbf{IC}(x) = J^\circ (\Phi_{S_x}^* \Phi_{T_x}^{+,*} e_x)$. The Identifiability Criterion clearly brings into play the promoted subspace T_{x_0} and the interaction between the restriction of Φ to T_{x_0} and S_{x_0} . It is a generalization of the irrepresentable condition that has been studied in the literature for some popular regularizers, including the ℓ^1 -norm [17], analysis- ℓ^1 [40], and ℓ^1 - ℓ^2 [1]. See Section 6 for a comprehensive discussion.

5.1. Noiseless Identifiability

We begin with the noiseless case, i.e. $w = 0$ in (1). It turns out that in such a setting, $\mathbf{IC}(x_0) < 1$ is a sufficient condition for identifiability without any other particular assumption on the bounded gauge J , such as piecewise regularity. By identifiability, we mean the fact that x_0 is the unique solution of $(\mathcal{P}_0(y))$.

Theorem 4. *Let $x_0 \in \mathbb{R}^N$ and $T = T_{x_0}$. We assume that (\mathcal{C}_T) holds and $\mathbf{IC}(x_0) < 1$. Then x_0 is the unique solution of $(\mathcal{P}_0(y))$.*

5.2. Exact Model Selection

It turns out that even in presence of noise in the measurements y according to (1), condition $\mathbf{IC}(x_0) < 1$ characterizes also those vectors where $(\mathcal{P}_\lambda(y))$ with piecewise regular gauges provides a robust selection of the model subspace of x_0 . Our main contribution is the following theorem.

Theorem 5. *Let $x_0 \in \mathbb{R}^N$ and $T = T_{x_0}$. We suppose that J is a piecewise regular gauge x_0 with the corresponding parameters $(\Gamma, \nu_{x_0}, \mu_{x_0}, \tau_{x_0}, \xi_{x_0})$. Assume that (\mathcal{C}_T) holds and $\mathbf{IC}(x_0) < 1$. Then there exist positive constants (A_T, B_T) that solely depend on T and a constant $C(x_0)$ such that if w and λ obey*

$$\frac{A_T}{1 - \mathbf{IC}(x_0)} \|w\| \leq \lambda \leq \nu_{x_0} \min(B_T, C(x_0)) \quad (7)$$

the solution x^* of $(\mathcal{P}_\lambda(y))$ with noisy measurements y is unique, and satisfies $T_{x^*} = T$. Furthermore, one has

$$\|x_0 - x^*\| = O\left(\max(\|w\|, \lambda)\right).$$

Clearly this result asserts that exact recovery of T_{x_0} from noisy partial measurements is possible with the proviso that the regularization parameter λ lies in the interval (7). The value λ should be large enough to reject noise, but small enough to recover the entire subspace T_{x_0} . In order for the constraint (7) to be non-empty, the noise-to-signal level $\|w\|/\nu_{x_0}$ should be small enough, i.e.

$$\frac{\|w\|}{\nu_{x_0}} \leq \frac{1 - \mathbf{IC}(x_0)}{A_T} \min(B_T, C(x_0)) .$$

The constant $C(x_0)$ involved in this bound depends on x_0 and has the form

$$C(x_0) = \frac{1 - \mathbf{IC}(x_0)}{\xi_{x_0} \nu_{x_0}} H\left(\frac{D_T \mu_{x_0} + \tau_{x_0}}{\xi_{x_0}}\right)$$

where $H(\beta) = \frac{\beta + 1/2}{E_T \beta} \varphi\left(\frac{2\beta}{(\beta + 1)^2}\right)$ and $\varphi(u) = \sqrt{1 + u} - 1$.

The constants (D_T, E_T) only depend on T . $C(x_0)$ captures the influence of the parameters $\pi_{x_0} = (\mu_{x_0}, \tau_{x_0}, \xi_{x_0})$, where the latter reflect the geometry of the regularizing gauge J at x_0 . More precisely, the larger $C(x_0)$, the more tolerant the recovery is to noise. Thus favorable regularizers are those where $C(x_0)$ is large, or equivalently where π_{x_0} has small entries, since H is a strictly decreasing function.

6. Examples of Piecewise Regular Gauges

6.1. Synthesis ℓ_1 Sparsity

The regularized problem $(\mathcal{P}_\lambda(y))$ with $J(x) = \|x\|_1 = \sum_{i=1}^N |x_i|$ promotes sparse solutions. It goes by the name of Lasso [38] in the statistical literature, and Basis Pursuit DeNoising (or Basis Pursuit in the noiseless case) [10] in signal processing.

The ℓ^1 norm is a PRG. The norm $J(x) = \|x\|_1$ is a symmetric (bounded) strong gauge. More precisely, we have the following result.

Proposition 11. $J = \|\cdot\|_1$ is a symmetric strong gauge with

$$T_x = \{\eta \in \mathbb{R}^N : \forall j \notin I, \eta_j = 0\}, \quad S_x = \{\eta \in \mathbb{R}^N : \forall i \in I, \eta_i = 0\}, \\ e_x = \text{sign}(x), \quad f_x = e_x, \quad J_{f_x}^\circ = \|\cdot\|_\infty + \iota_{S_x},$$

where $I = I(x) = \{i : x_i \neq 0\}$. Moreover, it is strongly piecewise regular with

$$\Gamma = \|\cdot\|_\infty, \quad \nu_x = (1 - \delta) \min_{i \in I} |x_i|, \delta \in]0, 1] \quad \text{and} \quad \mu_x = \tau_x = \xi_x = 0.$$

Relation to previous work. The theoretical recovery guarantees of ℓ^1 -regularization have been extensively studied in the recent years. There is of course a huge literature on the subject, and covering it comprehensively is beyond the scope of this paper. In this section, we restrict our overview to those works pertaining to ours, i.e., sparsity pattern recovery in presence of noise.

For instance, an identifiability criterion was introduced in [17]. Let $s \in \{-1, 0, +1\}^N$ and I its support. Suppose that $\Phi_{(I)}$ has full column rank, which is precisely (\mathcal{C}_T) in this case. The synthesis identifiability criterion \mathbf{IC}_{ℓ^1} of s is defined as

$$\mathbf{IC}_{\ell^1}(s) = \|\Phi_{(I^c)}^* \Phi_{(I)}^{+,*} s_{(I)}\|_\infty = \max_{j \in I^c} |\langle \Phi_j, \Phi_{(I)}^{+,*} s_{(I)} \rangle|.$$

From Definition 10 and Proposition 11, one immediately recognizes that $\mathbf{IC}_{\ell^1}(\text{sign}(x)) = \mathbf{IC}(x)$. The condition $\mathbf{IC}_{\ell^1}(\text{sign}(x)) < 1$, also known as the irrepresentable condition in the statistical literature, was proposed [17] for exact support (and sign) pattern recovery with ℓ_1 -regularization from partial noisy measurements. In this respect, this work can then be viewed as a special instance of ours, as Theorem 5 in this case ensures recovery of the support pattern.

6.2. Analysis ℓ^1 Sparsity

Let $D = (d_i)_{i=1}^P$ be a collection of P atoms $d_i \in \mathbb{R}^N$. The analysis semi-norm associated to D is $J(x) = \|D^*x\|_1 = \sum_{i=1}^P |\langle d_i, x \rangle|$. Obviously, the synthesis ℓ^1 -regularization corresponds to $D = \text{Id}$. Popular examples of analysis-type ℓ^1 semi-norms include for instance the discrete anisotropic total variation [35], the Fused Lasso [39] and shift invariant wavelets [36].

The analysis ℓ^1 semi-norm is a PRG. The semi-norm $J(x) = \|D^*x\|_1$ is a symmetric piecewise regular gauge. This is formalized in the following proposition whose proof is a straightforward application of Proposition 9, Proposition 10 and Proposition 11.

Proposition 12. $J = \|D^* \cdot\|_1$ is a symmetric (bounded) gauge with

$$\begin{aligned} T_x &= \text{Ker}(D_{(I^c)}^*) = \{\eta \in \mathbb{R}^N : \forall j \notin I, \langle d_j, \eta_j \rangle = 0\}, \quad S_x = \text{Im}(D_{I^c}), \\ e_x &= P_{\text{Ker}(D_{I^c}^*)} D \text{sign}(D^* x), \quad f_x = D \text{sign}(D^* x), \\ J_{f_x}^\circ(\eta) &= \inf_{z \in \text{Ker}(D_{(I^c)})} \|D_{(I^c)}^+ \eta + z\|_\infty, \quad \text{for } \eta \in S_x, \end{aligned}$$

where $I = I(x) = \{i : \langle d_i, x_i \rangle \neq 0\}$. Moreover, it is piecewise regular with parameters

$$\nu_x = (1 - \delta) \min_{i \in I} |\langle d_i, x_i \rangle|, \delta \in]0, 1] \quad \text{and} \quad \mu_x = \tau_x = \xi_x = 0.$$

Relation to previous work. Some insights on the relation and distinction between synthesis- and analysis-based sparsity regularizations were first given in [14]. When D is orthogonal, and more generally when D is square and invertible, the two forms of regularization are equivalent in the sense that the set of minimizers of one problem can be retrieved from that of an equivalent form of the other through a bijective change of variable. It is only recently that theoretical guarantees of ℓ^1 -analysis sparse regularization have been investigated, see [40] for a comprehensive review. Among such a work, the authors in [26] propose a null space property for identifiability in the noiseless case. The most relevant work to ours here is that of [40], where the authors prove exact robust recovery of the support and sign patterns under conditions that are a specialization of those in Theorem 5.

More precisely, let I be the support of $D^* x_0$, and s its sign vector. Denote $T = T_{x_0} = S^\perp = \text{Ker}(D_{I^c}^*)$, $e_{x_0} = \text{sign}(D^* x_0) = s$, $e = e_{x_0} = P_T D s$, $f = f_{x_0} = D s$. From Definition 10 and Proposition 12, the criterion $\mathbf{IC}(x_0)$ in this case takes the form

$$\begin{aligned} \mathbf{IC}(x_0) &= J_{f_x}^\circ(\Phi_S^* \Phi_T^{+,*} P_T D s - P_S D s) \\ &= \inf_{z \in \text{Ker}(D_{(I^c)})} \|D_{(I^c)}^+(\Phi_S^* \Phi_T^{+,*} P_T - P_S) D s + z\|_\infty \\ &= \inf_{z \in \text{Ker}(D_{(I^c)})} \|D_{(I^c)}^+((\text{Id} - P_T) \Phi^* \Phi P_T (\Phi_T^* \Phi_T)^{-1} P_T - P_S) D s + z\|_\infty \\ &= \inf_{z \in \text{Ker}(D_{(I^c)})} \|D_{(I^c)}^+(\Phi^* \Phi P_T (\Phi_T^* \Phi_T)^{-1} P_T - (P_T + P_S)) D s + z\|_\infty \\ &= \inf_{z \in \text{Ker}(D_{(I^c)})} \|D_{(I^c)}^+(\Phi^* \Phi P_T (\Phi_T^* \Phi_T)^{-1} P_T - \text{Id}) D_{(I)} s_{(I)} + z\|_\infty. \end{aligned}$$

Introducing U as a matrix whose columns form a basis of T , $\mathbf{IC}(x_0)$ can be equivalently rewritten

$$\mathbf{IC}(x_0) = \inf_{z \in \text{Ker}(D_{(I^c)})} \|D_{(I^c)}^+(\Phi^* \Phi A^{[I^c]} - \text{Id}) D_{(I)} s_{(I)} + z\|_\infty,$$

where $A^{[I^c]} = U(U^* \Phi^* \Phi U)^{-1} U^*$. We recover exactly the expression of the \mathbf{IC}_{ℓ^1-D} introduced in [40].

6.3. ℓ^∞ Antisparsity Regularization

Regularization by the ℓ^∞ -norm corresponds to taking $J(x) = \|x\|_\infty = \max_{1 \leq i \leq N} |x_i|$.

It plays a prominent role in a variety of applications including approximate nearest neighbor search [21] or vector quantization [24]; see also [37] and references therein.

The ℓ^∞ -norm is a PRG. The norm $J(x) = \|x\|_\infty$ is a symmetric piecewise regular gauge, but unlike the ℓ^1 -norm, it is not strongly so (except for $N = 2$). Therefore, in the following proposition, we rule out the trivial case $x = 0$.

Proposition 13. $J = \|\cdot\|_\infty$ is a symmetric (bounded) gauge with

$$S_x = \{\eta : \eta_{(I^c)} = 0 \text{ and } \langle \eta_{(I)}, s_{(I)} \rangle = 0\}, \quad T_x = \{\alpha : \alpha_{(I)} = \rho s_{(I)} \text{ for } \rho \in \mathbb{R}\},$$

$$e_x = \frac{s}{|I|}, \quad f_x = e_x, \quad J_{f_x}^\circ(\eta) = \max_{i \in I} (-|I|s_i \eta_i)_+ \text{ for } \eta \in S_x,$$

where $s = \text{sign}(x)$ and $I = I(x) = \{i : |x_i| = \|x\|_\infty\}$. Moreover, it is piecewise regular with

$$\Gamma = \|\cdot\|_1, \quad \nu_x = (1 - \delta)(\|x\|_\infty - \max_{j \notin I} |x_j|), \delta \in]0, 1] \quad \text{and} \quad \mu_x = \tau_x = \xi_x = 0.$$

Relation to previous work. In the noiseless case, i.e. $(\mathcal{P}_0(y))$ with $J = \|\cdot\|_\infty$, theoretical analysis of ℓ^∞ -regularization goes back to the 70's through the work of [6]. [24] provided results that characterize signal representations with small (but not necessarily minimal) ℓ^∞ -norm subject to linear constraints. A necessary and sufficient condition for a vector to be the unique minimizer of $(\mathcal{P}_0(y))$ is derived in [25]. The work of [13] analyzes recovery guarantees by ℓ^∞ -regularization in a noiseless random sensing setting.

The authors in [37] analyzed the properties of solutions obtained from a constrained form of $(\mathcal{P}_\lambda(y))$ with $J = \|\cdot\|_\infty$. In particular, they improved and generalized the bound of [24] on the ℓ^∞ of the solution.

The work of [3, 28] studies robust recovery with regularization using a subclass of polyhedral norms obtained by convex relaxation of combinatorial penalties. Although this covers the case of the ℓ^∞ -norm, their notion of support is however, completely different from ours. We will come back to this work with a more detailed discussion in Section 6.5.

6.4. Block/Group Sparsity Regularization

Let's recall from Section 2.2 that \mathcal{B} is a uniform disjoint partition of $\{1, \dots, N\}$,

$$\{1, \dots, N\} = \bigcup_{b \in \mathcal{B}} b, \quad b \cap b' = \emptyset, \quad \forall b \neq b'.$$

The $\ell^1 - \ell^2$ norm of x is

$$J(x) = \|x\|_{\mathcal{B}} = \sum_{b \in \mathcal{B}} \|x_b\|.$$

This prior has been advocated when the signal exhibits a structured sparsity pattern where the entries are assumed to be clustered in few non-zero groups; see for instance [4, 44]. The corresponding regularized problem ($\mathcal{P}_\lambda(y)$) is known as the group Lasso.

The ℓ^1 - ℓ^2 norm is a PRG. The $\ell^1 - \ell^2$ norm is a symmetric strongly piecewise regular gauge.

Proposition 14. *The $\ell^1 - \ell^2$ norm associated to the partition \mathcal{B} is a symmetric (bounded) strong gauge with*

$$T_x = \{\eta : \forall j \notin I, \eta_j = 0\}, \quad S_x = \{\eta : \forall i \in I, \eta_i = 0\}, \\ e_x = (\mathcal{N}(x_b))_{b \in \mathcal{B}}, \quad f_x = e_x, \quad J^\circ = \|\cdot\|_{\infty,2} + \iota_{S_x},$$

where $I = I(x) = \{b : x_b \neq 0\}$, and $\mathcal{N}(a) = a/\|a\|$ if $a \neq 0$, and $\mathcal{N}(0) = 0$. Moreover, it is strongly piecewise regular with

$$\Gamma = \|\cdot\|_{\infty,2}, \quad \nu_x = (1 - \delta) \min_{b \in I} \|x_b\|, \delta \in]0, 1[\quad \mu_x = \frac{\sqrt{2}}{\nu_x} \quad \text{and} \quad \tau_x = \xi_x = 0.$$

Relation to previous work. Theoretical guarantees of the group Lasso have been investigated by several authors under different performance criteria; see e.g. [44, 34, 1, 11, 23, 43] to cite only a few. In particular, the author in [1] studies the asymptotic group selection consistency of the group Lasso in the overdetermined case, under a group irrepresentable condition. This condition also appears in noiseless identifiability in the work of [7]. The group irrepresentable condition is nothing but the specialization to the group Lasso of our condition based on $\mathbf{IC}(x_0)$. Indeed, using Definition 10 and Proposition 14, and assuming that $\Phi_{(I)}$ is full column rank (i.e. (\mathcal{C}_T) is fulfilled), $\mathbf{IC}(x_0)$ reads

$$\mathbf{IC}(x_0) = \left\| \Phi_{(I^c)}^* \Phi_{(I)}^{+,*} \left(\frac{x_b}{\|x_b\|} \right)_{b \in I} \right\|_{\infty,2}. \quad (8)$$

It is worth mentioning that the discrete isotropic total variation in d -dimension, $d \geq 2$, can be viewed as an analysis-type $\ell^1 - \ell^2$ semi-norm. Piecewise regularity and theoretical recovery guarantees with such a regularization can be retrieved from those of this paper using the results on the pre-composition rule given in Section 4.2.2.

6.5. Polyhedral Regularization

A particular case of analysis priors are polyhedral gauges, that are in general not strongly piecewise regular. The ℓ^1 and ℓ^∞ norms are special cases of polyhedral priors (ℓ^1 being strongly piecewise regular, while ℓ^∞ being only piecewise regular). There are two alternative ways to define a polyhedral gauge. The H -representation encodes the gauge through the hyperplanes that support the polygonal facets of its unit level set. The V -representation encodes the gauge through the vertices that are the extreme points of this unit level set. We focus here on the H -representation.

A polyhedral gauge is a PRG. A polyhedral gauge in the H -representation is defined as

$$J(x) = \max_{1 \leq i \leq N_H} (\langle x, h_i \rangle)_+ = J_0(H^*x) \quad \text{where} \quad J_0(u) = \max_{1 \leq i \leq N_H} (u_i)_+,$$

and we have defined $H = (h_i)_{i=1}^{N_H} \in \mathbb{R}^{N \times N_H}$.

Such a polyhedral gauge can also be thought as an analysis gauge as considered in Section 4.2.2 by identifying $D = H$. One can then characterize decomposability and piecewise regularity of J_0 and then invoke Proposition 9 and 10 to derive those of J . This is what we are about to do. In the following, we denote $(a^i)_{1 \leq i \leq N_H}$ the standard basis of \mathbb{R}^{N_H} .

Proposition 15. $J_0(u) = \max_{1 \leq i \leq N_H} (u_i)_+$ is a (bounded) gauge and,

- If $u_i \leq 0, \forall i \in \{1, \dots, N_H\}$, then

$$\begin{aligned} S_u &= \text{span} (a^i)_{i \in I_0}, \quad T_u = \text{span} (a^i)_{i \notin I_0}, \\ e_u &= 0, \quad f_u = \mu \sum_{i \in I_0} a^i, \quad \text{for any } 0 < \mu < 1, \\ J_{f_u}^\circ(\eta) &= \inf_{\tau \geq \max_{i \in I_0} (-\eta_i)_+ / \mu} \max \left(\tau \mu |I_0| + \sum_{i \in I_0} \eta_i, \tau \right) \quad \text{for } \eta \in S_u, \end{aligned}$$

where

$$I_0 = \{i \in \{1, \dots, N_H\} : u_i = J_0(u) = 0\}.$$

- If $\exists i \in \{1, \dots, N_H\}$ such that $u_i > 0$, then

$$\begin{aligned} S_u &= \left\{ \eta : \eta_{(I_+^c)} = 0 \quad \text{and} \quad \langle \eta_{(I_+)}, s_{(I_+)} \rangle = 0 \right\}, \\ T_u &= \left\{ \alpha : \alpha_{(I_+)} = \mu s_{(I_+)} \quad \text{for } \mu \in \mathbb{R} \right\}, \\ e_u &= \frac{s}{|I_+|}, \quad f_u = e_u, \quad J_{f_u}^\circ(\eta) = \max_{i \in I_+} (-|I_+| \eta_i)_+ \quad \text{for } \eta \in S_u, \end{aligned}$$

where

$$s = \sum_{i \in I_+} a^i \quad \text{and} \quad I_+ = \{i \in \{1, \dots, N_H\} : u_i = J_0(u) \quad \text{and} \quad u_i > 0\}.$$

Moreover, it is piecewise regular with parameters (assuming $I_+ \neq \emptyset$)

$$\nu_u = (1 - \delta) \left(\max_{i \in I_+} u_i - \max_{j \notin I_+, u_j > 0} u_j \right), \delta \in]0, 1] \quad \text{and} \quad \mu_u = \tau_u = \xi_u = 0.$$

Relation to previous work. As stated in the case of ℓ^∞ -norm, the work of [3] considers robust recovery with a subclass of polyhedral norms but his notion of support is different from ours. The work [30] studies numerically some polyhedral regularizations. Again in a compressed sensing scenario, the work of [9] studies a subset of polyhedral

regularizations to get sharp estimates of the number of measurements for exact and ℓ_2 -stable recovery. The closest work to ours is that reported in [41], where theoretical recovery guarantees by polyhedral regularization were provided under similar conditions to ours and with the same notion of support as considered above. However only bounded coercive polyhedral gauges were considered there.

6.6. Nuclear Norm Regularization

The nuclear norm, or trace norm, has been proposed as an effective convex relaxation for rank minimization problems to provably recover low-rank matrices [8, 32]. The nuclear norm of a $N_1 \times N_2$ rank- r matrix x is defined as $J(x) = \|x\|_* = \|\text{diag}(\Sigma)\|_1$, where $x = U\Sigma V^*$ is a reduced singular value decomposition (SVD) of x , with $\text{diag}(\Sigma) = (\sigma_i)_{1 \leq i \leq r}$ the singular values, U ($N_1 \times r$) and V ($N_2 \times r$) are the matrices of left and right singular vectors.

The nuclear norm is a PRG. In the following, we show that the nuclear norm is a symmetric strongly piecewise regular gauge.

Proposition 16. *The nuclear norm is a symmetric strong gauge with*

$$S_x = \left\{ U_\perp^* C V_\perp : C \in \mathbb{R}^{(N_1-r) \times (N_2-r)} \right\},$$

$$T_x = \left\{ U A^* + B V^* : A \in \mathbb{R}^{N_2 \times r}, B \in \mathbb{R}^{N_1 \times r} \right\} = \left\{ Z \in \mathbb{R}^{(N_1-r) \times (N_2-r)} : U_\perp^* Z V_\perp = 0 \right\},$$

$$e_x = UV^*, \quad f_x = e_x, \quad J^\circ(x) = \max_i \sigma_i + \iota_{S_x},$$

where U_\perp, V_\perp span the orthogonal of the ranges of U, V . Moreover, it is strongly piecewise regular with parameters

$$\nu_x = \frac{1}{4} \min_{i \in I} |\sigma_i|, \quad \mu_x = \frac{1}{\nu_x} \quad \text{and} \quad \tau_x = \xi_x = 0.$$

It can be observed that $\dim(T_x) = r(N_1 + N_2 - r)$ and $\dim(S_x) = N_1 N_2 - \dim(T_x) = N_1 N_2 - r(N_1 + N_2) + r^2$.

Relation to previous work. The work in [2] defines an irrepresentability condition for the nuclear norm. This irrepresentability condition is assumed in [2] to establish asymptotic rank consistency of nuclear norm minimization in the overdetermined case. Similarly to (8), this condition can be shown to be equivalent to our condition $\mathbf{IC}(x_0) < 1$, when specialized to the case of the nuclear norm. Note also that [2] provides a proof of the PRG property of the nuclear norm (Proposition 16). In their unified framework, the work [32] derives tight bounds on the number of Gaussian measurements needed for exact and robust recovery by nuclear norm minimization. The work of [32] complements these results, using the irrepresentable condition tailored to the nuclear norm, to get tight bounds for noiseless identifiability.

Let us also point out that the recent work of [19] considers an analysis-type form of the nuclear norm, the so-called trace Lasso. However no recovery guarantees (with or

without noise) were given. In fact, combining Proposition 16 and the results on the pre-composition rule given in Section 4.2.2, one can establish piecewise regularity and derive theoretical recovery guarantees for the trace Lasso.

7. Conclusion

In this paper, we introduced the notion of piecewise regular gauge as a generic convex regularization framework, and presented a unified view of to derive exact and robust recovery guarantees for a large class of convex regularizations. In particular, we provided sufficient conditions ensuring uniqueness of the minimizer to both $(\mathcal{P}_\lambda(y))$ and $(\mathcal{P}_0(y))$, whose byproduct is to guarantee exact recovery of the original object x_0 in the noiseless case by solving $(\mathcal{P}_0(y))$. In presence of noise, sufficient conditions were given to certify exact recovery of the model subspace underlying x_0 . As shown in the considered examples, these results encompass a variety of cases extensively studied in the literature (e.g. ℓ^1 , analysis ℓ^1 , $\ell^1 - \ell^2$, nuclear norm), as well as less popular ones (inf, polyhedral).

Appendices

A. Proofs of Section 2

Proof of Lemma 1. (i)-(iii) are obtained from [20, Theorem V.1.2.5]. (iv) is obtained by combining [20, Corollary V.1.2.6 and Proposition IV.3.2.5]. (v): the second statement follows by combining (iii)-(iv), while the first part is the second one written in $\text{dom } \gamma_C = \text{aff } C = \text{span } C$ since $0 \in C$. \square

Proof of Lemma 2. (i) follows from [33, Theorem 15.1]. (ii) [33, Corollary 15.1.1] or [20, Proposition V.3.2.4]. (iii) [33, Corollary 15.1.2] or [20, Proposition V.3.2.5]. \square

Proof of Lemma 3. We have from Lemma 2 and calculus rules on support functions,

$$\gamma_{(C_1+C_2)^\circ} = \sigma_{C_1+C_2} = \sigma_{C_1} + \sigma_{C_2} .$$

Thus

$$(C_1 + C_2)^\circ = \{u : \sigma_{C_1}(u) + \sigma_{C_2}(u) \leq 1\} .$$

This yields that

$$\begin{aligned}
\gamma_{C_1+C_2}(x) &= \sigma_{(C_1+C_2)^\circ}(x) \\
&= \sigma_{\sigma_{C_1}(u)+\sigma_{C_2}(u)\leq 1} \\
&= \sup_{\sigma_{C_1}(u)+\sigma_{C_2}(u)\leq 1} \langle u, x \rangle \\
&= \sup_{\rho\in[0,1]} \sup_{\sigma_{C_1}(u)\leq\rho, \sigma_{C_2}(u)\leq 1-\rho} \langle u, x \rangle \\
&= \sup_{\rho\in[0,1]} \sigma_{\sigma_{C_1}(u)\leq\rho} \overset{+}{\vee} \sigma_{\sigma_{C_2}(u)\leq 1-\rho}(x) \\
&= \sup_{\rho\in[0,1]} \rho\sigma_{\sigma_{C_1}(u)\leq 1} \overset{+}{\vee} (1-\rho)\sigma_{\sigma_{C_2}(u)\leq 1}(x) \\
&= \sup_{\rho\in[0,1]} \rho\sigma_{C_1^\circ} \overset{+}{\vee} (1-\rho)\sigma_{C_2^\circ}(x) \\
&= \sup_{\rho\in[0,1]} \sigma_{\rho C_1^\circ} \overset{+}{\vee} \sigma_{(1-\rho)C_2^\circ}(x) \\
&= \sup_{\rho\in[0,1]} \rho\gamma_{C_1} \overset{+}{\vee} (1-\rho)\gamma_{C_2}(x) ,
\end{aligned}$$

which is the first assertion.

The last identity can be rewritten

$$\gamma_{C_1+C_2}(x) = \sup_{\rho\in[0,1]} \inf_{x_1+x_2=x} \rho\gamma_{C_1}(x_1) + (1-\rho)\gamma_{C_2}(x_2) .$$

Under the boundedness and continuity assumption of the lemma, the objective in the sup inf is a continuous bounded concave-convex function on $[0, 1]\times\{(x_1, x_2) : x_1 + x_2 = x\}$. Since the latter sets are non-empty, closed and convex, and $[0, 1]$ is obviously bounded, we have from using [33, Corollary 37.3.2]

$$\begin{aligned}
\gamma_{C_1+C_2}(x) &= \inf_{z\in\mathbb{R}^N} \sup_{\rho\in[0,1]} \rho\gamma_{C_1}(z) + (1-\rho)\gamma_{C_2}(x-z) \\
&= \inf_{z\in\mathbb{R}^N} \max(\gamma_{C_1}(z), \gamma_{C_2}(x-z)) .
\end{aligned}$$

□

Proof of Lemma 4. It is immediate to see that $D(C)$ is a closed convex set containing the origin. Moreover, we have $\text{Im}(D^*) \cap \text{dom}(\sigma_C) \neq \emptyset$, since the origin is in both of them. Thus, using [20, Theorem X.2.1.1] and Lemma 2, we have

$$\gamma_{(D(C))^\circ} = \sigma_{D(C)} = (\iota_{D(C)})^* = \sigma_C \circ D^* .$$

Now, as by assumption $0 \in \text{ri } C$, we have $0 \in \text{ri}(C^\circ)$, and therefore $\text{Im}(D^*) \cap \text{ri}(C^\circ) \neq \emptyset$.

By virtue of [20, Theorem X.2.2.3] and Lemma 2, we get

$$\begin{aligned}
\gamma_{D(C)}(x) &= \sigma_{(D(C))^\circ}(x) \\
&= \sigma_{\sigma_C \circ D^*(u) \leq 1}(x) \\
&= (\iota_{\sigma_C(w) \leq 1} \circ D^*)^*(x) \\
&= \inf_v \sigma_{\sigma_C(w) \leq 1}(v) \quad \text{s.t.} \quad Dv = x \\
&= \inf_{z \in \text{Ker}(D)} \sigma_{\sigma_C(w) \leq 1}(D^+x + z) \\
&= \inf_{z \in \text{Ker}(D)} \sigma_{\sigma_C(w) \leq 1}(D^+x + z) \\
&= \inf_{z \in \text{Ker}(D)} \gamma_C(D^+x + z) .
\end{aligned}$$

□

Proof of Proposition 1.

- (i) Each element of \bar{S}_x can be written as $u = \sum_{i=1}^k \rho_i \eta_i$, for $k > 0$, where $\eta_i \in \partial J(x)$ and $\sum_{i=1}^k \rho_i = 1$. By Fenchel identity applied to the gauge J , and using Lemma 2(iii), we have

$$\langle x, \eta_i \rangle = J(x) + \iota_{C^\circ}(\eta_i), \quad \forall i .$$

Since $\eta_i \in \partial J(x) \subseteq C^\circ$, we get

$$\langle x, \eta_i \rangle = J(x), \quad \forall i ,$$

Multiplying by ρ_i and summing this identity over i and using the fact that $\sum_{i=1}^k \rho_i = 1$ we obtain the desired result.

- (ii) This is due to the fact that e_x is the orthogonal projection of 0 on the affine space \bar{S}_x . It is therefore an element of $\bar{S}_x \cap (\bar{S}_x - e_x)^\perp$, i.e. $e_x \in \bar{S}_x \cap T_x$.
- (iii) This is straightforward from the fact that $S_x = \{\eta \in \mathbb{R}^N : \eta_{T_x} = 0\}$, $\bar{S}_x = S_x + e_x$ and $e_x \in T_x$ from (ii).
- (iv) For any $v \in S_x$, we have $v + e_x \in \bar{S}_x$ since $e_x \in \bar{S}_x$. Thus applying (i), we get $\langle x, e_x + v \rangle = J(x)$ and $\langle x, e_x \rangle = J(x)$. Combining both identities implies that $\langle x, v \rangle = 0$, $\forall v \in S_x$, or equivalently that $x \in S_x^\perp = T_x$.

□

Proof of Proposition 2. Let $x \in \mathbb{R}^N$. We have

$$\partial J(x) = \mathbf{F}_{C^\circ}(x) = H \cap C^\circ ,$$

where $H = \{\eta \in \mathbb{R}^N : \langle \eta, x \rangle = J(x)\}$ is the supporting hyperplane of C° at x . By Proposition 1(i), we have

$$\bar{S}_x = \text{aff } \partial J(x) \subseteq H ,$$

which implies that

$$\bar{S}_x \cap C^\circ \subseteq H \cap C^\circ.$$

The converse inclusion is true since $\partial J(x) = H \cap C^\circ \subseteq \bar{S}_x$. \square

Proof of Proposition 3. The first assertion follows from Lemma 1(v) since $0 \in \text{ri}(\partial J(x) - f_x)$. Let's now turn to the second part. Since $f_x \in \text{ri} \partial J(x) \subset \bar{S}_x$, Proposition 1 implies that $f_x = P_{S_x}(f_x) + P_{T_x}(f_x) = P_{S_x}(f_x) + e_x$. Hence, using Proposition 2, we get

$$\begin{aligned} \partial J(x) - f_x &= (C^\circ - f_x) \cap (\bar{S}_x - f_x) \\ &= (C^\circ - f_x) \cap (S_x - \{P_{S_x}(f_x)\}) \\ &= (C^\circ - f_x) \cap S_x. \end{aligned}$$

We therefore obtain

$$\begin{aligned} J_f^\circ(\eta) &= \gamma_{(C^\circ - f_x) \cap S_x}(\eta) \\ &= \max(\gamma_{C^\circ - f_x}(\eta), \gamma_{S_x}(\eta)) \\ &= \max(\gamma_{C^\circ - f_x}(\eta), \iota_{S_x}(\eta)) \\ &= \gamma_{C^\circ - f_x}(\eta) + \iota_{S_x}(\eta). \end{aligned}$$

At this stage, Lemma 3 does not apply straightforwardly since $0 \in C^\circ$ but $f_x \neq 0$ in general. However, proceeding as in the proof of that lemma, we arrive at

$$\gamma_{C^\circ + \{-f_x\}}(\eta) = \sup_{\rho \in [0,1]} \rho J^\circ \overset{+}{\vee} (1 - \rho) \sigma_{\{-f_x\}^\circ}(\eta)$$

where, from Definition 2, $\{-f_x\}^\circ = \{\eta : \langle \eta, f_x \rangle \geq -1\}$, which indeed contains the origin as an interior point. Continuing from the last equality, we get

$$\begin{aligned} \gamma_{C^\circ + \{-f_x\}}(\eta) &= \sup_{\rho \in [0,1]} \rho J^\circ \overset{+}{\vee} (1 - \rho) \gamma_{\{-f_x\}^{\circ\circ}}(\eta) \\ &= \sup_{\rho \in [0,1]} \rho J^\circ \overset{+}{\vee} (1 - \rho) \gamma_{\overline{\text{co}}(\{-f_x\} \cup \{0\})}(\eta) \\ &= \sup_{\rho \in [0,1]} \rho J^\circ \overset{+}{\vee} (1 - \rho) \gamma_{\{-\mu f_x : \mu \in [0,1]\}}(\eta). \end{aligned}$$

It is easy to see that

$$\gamma_{\{-\mu f_x : \mu \in [0,1]\}}(-\eta) = \begin{cases} \tau & \text{if } \eta \in \tau f_x, \tau \in \mathbb{R}_+, \\ +\infty & \text{otherwise.} \end{cases}$$

Thus

$$\gamma_{C^\circ + \{-f_x\}}(\eta) = \sup_{\rho \in [0,1]} \inf_{\tau \geq 0} \rho J^\circ(\tau f_x + \eta) + (1 - \rho)\tau.$$

Recalling that J° is a bounded gauge, hence continuous, the objective in the supinf fulfills the assumption of the second assertion of Lemma 3, whence we get

$$\gamma_{C^\circ + \{-f_x\}}(\eta) = \inf_{\tau \geq 0} \max(J^\circ(\tau f_x + \eta), \tau) .$$

□

Proof of Proposition 4. The gauge J_{f_x} is the support function of the set

$$\mathcal{K}_x \stackrel{\text{def.}}{=} \partial J(x) - f_x = \{\eta \in \mathbb{R}^N : J_{f_x}^\circ(\eta) \leq 1\} \subset S_x ,$$

where the inclusion follows from Proposition 3.

- (i) Since \mathcal{K}_x is a bounded set, its support function is also bounded [20, Proposition V.2.1.3]. It follows that $\text{dom } J_{f_x} = \mathbb{R}^N$.
- (ii) We have

$$\begin{aligned} J_{f_x}(d) &= \sup_{\eta \in \mathcal{K}_x} \langle \eta, d \rangle = \sup_{J_{f_x}^\circ(\eta) \leq 1} \langle \eta, d \rangle = \sup_{J_{f_x}^\circ(\eta_{S_x}) \leq 1} \langle \eta_{S_x}, d \rangle \\ &= \sup_{\eta \in \mathcal{K}_x} \langle \eta, d_{T_x} \rangle + \langle \eta, d_{S_x} \rangle = \sup_{\eta \in \mathcal{K}_x} \langle \eta, d_{S_x} \rangle \\ &= J_{f_x}(d_{S_x}) , \end{aligned}$$

where we used the fact that $\langle \eta, d_{T_x} \rangle = 0$ on \mathcal{K}_x .

- (iii) As a consequence of (ii), $J_{f_x}(d_{T_x}) = 0$. Clearly, $T_x \subset \text{Ker}(J_{f_x})$ and J_{f_x} is constant along all affine subspaces parallel to T_x . But, since $0 \in \text{ri } \mathcal{K}_x$, excluding the origin, the supremum in J_{f_x} is always attained at some nonzero $\eta \in \mathcal{K}_x \subset S_x$. Thus $J_{f_x}(d) > 0$ for all d such that $d \notin T_x$. This shows that actually $\text{Ker}(J_{f_x}) = T_x$. In particular, this yields that on S_x , the gauge J_{f_x} is coercive.
- (iv) Using some calculus rules with support functions and assertion (ii), we have

$$\begin{aligned} J_{f_x}(d) &= J_{f_x}(d_{S_x}) = \sigma_{(C^\circ + \{-f_x\}) \cap S_x}(d_{S_x}) \\ &= \overline{\text{co}}(\inf(\sigma_{C^\circ + \{-f_x\}}(d_{S_x}), \sigma_S(d_{S_x}))) \\ &= \overline{\text{co}}(\inf(\sigma_{C^\circ + \{-f_x\}}(d_{S_x}), \iota_T(d_{S_x}))) \\ &= \sigma_{C^\circ + \{-f_x\}}(d_{S_x}) \\ &= \sigma_{C^\circ}(d_{S_x}) - \langle \mathbb{P}_{S_x}(f_x), d_{S_x} \rangle \\ &= J(d_{S_x}) - \langle \mathbb{P}_{S_x}(f_x), d_{S_x} \rangle . \end{aligned}$$

□

Proof of Theorem 1. We embark from

$$\partial J(x) = \{\eta \in \mathbb{R}^N : \eta \in C^\circ \quad \text{and} \quad \langle \eta, x \rangle = J(x)\} .$$

Invoking Proposition 1, we get that for every $\eta \in \partial J(x)$, $\eta_{T_x} = e_x$, and $P_{T_x}(f_x) = e_x$. It remains now to uniquely characterize the part of the subdifferential lying in S_x , i.e. $\partial J(x) - e_x$. Since $f_x \in \text{ri } \partial J(x)$, we have from the one-to-one correspondence of Lemma 1(i) and the definition of the antipromoting gauge,

$$\begin{aligned} \eta \in \{ \eta \in \mathbb{R}^N : J_{f_x}^\circ(\eta_{S_x} - P_{S_x}(f_x)) \leq 1 \} &\iff \eta_{S_x} - P_{S_x}(f_x) \in \partial J(x) - f_x \\ &\iff \eta_{S_x} \in \partial J(x) - e_x \\ &\iff \eta \in \partial J(x) . \end{aligned}$$

□

Proof of Proposition 5. This is a convenient rewriting of the fact that x is a global minimizer if, and only if, 0 is a subgradient of the objective function at x .

(i) For problem $(\mathcal{P}_\lambda(y))$, this is equivalent to

$$\frac{1}{\lambda} \Phi^*(y - \Phi x) \in \partial J(x).$$

Projecting this relation on T and S yields the desired result.

(ii) Let's turn to problem $(\mathcal{P}_0(y))$. We have at any global minimizer x

$$0 \in \partial J(x) + \Phi^* N_{\{\alpha: \alpha=y\}}(\Phi x)$$

where $N_{\{\alpha: \alpha=y\}}(x)$ is the normal cone of the constraint set $\{\alpha : \alpha = y\}$ at x , which is obviously the whole space \mathbb{R}^Q . Thus, this monotone inclusion is equivalent to the existence of $\alpha \in \mathbb{R}^Q$ such that

$$\Phi^* \alpha \in \partial J(x) .$$

Projecting again this on T and S proves the assertion.

□

Proof of Lemma 5. Let $J = \gamma_C$, $x \in T$ and $x' \in S$.

\Rightarrow : By virtue of Lemma 2, we have

$$\begin{aligned}
J^\circ(x + x') &= \sup_{u \in C} \langle x + x', u \rangle \\
&= \sup_{J(u) \leq 1} \langle x + x', u \rangle \\
&= \sup J(u_T + u_S) \leq 1 \langle x, u_T \rangle + \langle x', u_S \rangle \\
&= \sup J(u_T) + J(u_S) \leq 1 \langle x, u_T \rangle + \langle x', u_S \rangle \\
&= \sup_{\rho \in [0,1]} \sup_{J(u_T) \leq \rho, J(u_S) \leq 1-\rho} \langle x, u_T \rangle + \langle x', u_S \rangle \\
&= \sup_{\rho \in [0,1]} \rho \sup_{J(u_T) \leq 1} \langle x, u_T \rangle + (1 - \rho) \sup_{J(u_S) \leq 1} \langle x', u_S \rangle \\
&= \sup_{\rho \in [0,1]} \rho \sup_{v \in C \cap T} \langle x, v \rangle + (1 - \rho) \sup_{w \in C \cap S} \langle x', w \rangle \\
&= \sup_{\rho \in [0,1]} \rho \sigma_{C \cap T}(x) + (1 - \rho) \sigma_{C \cap S}(x') \\
&= \max(\sigma_{C \cap T}(x), \sigma_{C \cap S}(x')) .
\end{aligned}$$

Since

$$\sigma_{C \cap T}(x) = \overline{\text{co}}(\inf(\sigma_C(x), \iota_S(x))) = \sigma_C(x) = J^\circ(x)$$

and

$$\sigma_{C \cap S}(x') = \overline{\text{co}}(\inf(\sigma_C(x'), \iota_T(x'))) = \sigma_C(x') = J^\circ(x') ,$$

the implication follows.

\Leftarrow : Using again Lemma 2, we get

$$\begin{aligned}
J(x + x') &= \sup_{u \in C^\circ} \langle x + x', u \rangle \\
&= \sup_{J^\circ(u_T + u_S) \leq 1} \langle x, u_T \rangle + \langle x', u_S \rangle \\
&= \sup_{\max(J^\circ(u_T), J^\circ(u_S)) \leq 1} \langle x, u_T \rangle + \langle x', u_S \rangle \\
&= \sup_{J^\circ(u_T) \leq 1, J^\circ(u_S) \leq 1} \langle x, u_T \rangle + \langle x', u_S \rangle \\
&= \sup_{v \in C^\circ \cap T} \langle x, v \rangle + \sup_{w \in C^\circ \cap S} \langle x', w \rangle \\
&= \sigma_{C^\circ \cap T}(x) + \sigma_{C^\circ \cap S}(x') \\
&= \overline{\text{co}}(\inf(\sigma_{C^\circ}(x), \iota_S(x))) + \overline{\text{co}}(\inf(\sigma_{C^\circ}(x'), \iota_T(x'))) \\
&= \sigma_{C^\circ}(x) + \sigma_{C^\circ}(x') \\
&= J(x) + J(x') .
\end{aligned}$$

This concludes the proof. \square

Proof of Proposition 6. Let $J = \gamma_C$. We only need to show that $J_{e_x}^\circ(\eta_{S_x}) = J^\circ(\eta_{S_x})$.

This follows from Proposition 3, Lemma 5 and Lemma 2(ii). Indeed,

$$\begin{aligned}
J_{e_x}^\circ(\eta_{S_x}) &= \inf_{\tau \geq 0} \max(J^\circ(\tau e_x + \eta_{S_x}), \tau) && \text{from Proposition 3,} \\
&= \inf_{\tau \geq 0} \max(\tau J^\circ(e_x), J^\circ(\eta_{S_x}), \tau) && \text{from Lemma 5,} \\
&= \inf_{\tau \geq 0} \max(J^\circ(\eta_{S_x}), \tau) && \text{from } e_x \in \partial J(x) \subset C^\circ, \\
&= J^\circ(\eta_{S_x}) .
\end{aligned}$$

□

B. Proofs of Section 3

Proof of Lemma 6. Let x_1, x_2 be two (global) minimizers of $(\mathcal{P}_\lambda(y))$. Suppose that $\Phi x^1 \neq \Phi x^2$. Define $x_t = tx_1 + (1-t)x_2$ for any $t \in (0, 1)$. By strict convexity of $u \mapsto \|y - u\|_2^2$, one has

$$\frac{1}{2}\|y - \Phi x_t\|_2^2 < \frac{t}{2}\|y - \Phi x_1\|_2^2 + \frac{1-t}{2}\|y - \Phi x_2\|_2^2.$$

Since J is convex, we get

$$J(x_t) \leq tJ(x_1) + (1-t)J(x_2).$$

Combining these two inequalities contradicts the fact that x_1, x_2 are global minimizers of $(\mathcal{P}_\lambda(y))$. □

Proof of Theorem 2. To prove this theorem, we need the following lemmata.

Lemma 7. *Let C be a non-empty closed convex set and f a proper lsc convex function. Let x be a minimizer of $\min_{z \in C} f(z)$. If*

$$f'(x, z - x) > 0 \quad \forall z \in C, z \neq x ,$$

then, x is the unique solution of f on C .

Proof. We first show that $t \mapsto (f(x + t(z - x)) - f(z))/t$ is non-decreasing on $(0, 1]$. Indeed, let $g : [0, 1] \rightarrow \mathbb{R}$ a convex function such that $g(0) = 0$. Let $(t, s) \in (0, 1]^2$ with $s > t$. Then,

$$\begin{aligned}
g(t) &= g(s(t/s)) = g(s(t/s) + (1-t/s)0) \\
&\leq t \frac{g(s)}{s} + (1-t/s)g(0) \\
&= t \frac{g(s)}{s} ,
\end{aligned}$$

which proves that $t \in (0, 1] \mapsto \frac{g(t)}{t}$ is non-decreasing on $(0, 1]$. Since f is convex, applying this result shows that the function

$$t \in (0, 1] \mapsto g(t) = f(x + t(z - x)) - f(z)$$

is such that $g(0) = 0$ and $g(t)/t$ is non-decreasing.

Assume now that that $f'(x, z - x) > 0$. Then, for every $x \in C$,

$$g(1) = f(z) - f(x) \geq f'(x, z - x) > 0, \quad \forall z \in C, z \neq x,$$

which is equivalent to x being the unique minimizer of f on C . \square

We now compute the directional derivative of a bounded gauge J .

Lemma 8. *The directional derivative $J'(x, \delta)$ at point $x \in \mathbb{R}^N$ in the direction δ reads*

$$J'(x, \delta) = \langle e_x, \delta_{T_x} \rangle + \langle P_{S_x}(f_x), \delta_{S_x} \rangle + J_{f_x}(\delta_{S_x}).$$

Proof. This comes directly from the structure of J_{f_x} . Indeed, one has

$$\begin{aligned} J_{f_x}(\delta_{S_x}) &= J_{f_x}(\delta) && \text{Using Proposition 4(ii)} \\ &= \sup_{\eta \in \partial J(x) - \{f_x\}} \langle \eta, \delta \rangle \\ &= -\langle \delta, f_x \rangle + \sup_{\eta \in \partial J(x)} \langle \eta, \delta \rangle \\ &= -\langle \delta, f_x \rangle + J'(x, \delta) \\ &= -\langle e_x, \delta_{T_x} \rangle - \langle P_{S_x}(f_x), \delta_{S_x} \rangle + J'(x, \delta). \end{aligned}$$

\square

We are now in position to show Theorem 3. We provide the proof for $(\mathcal{P}_\lambda(y))$. That of $(\mathcal{P}_0(y))$ is similar.

Let x be a solution of $(\mathcal{P}_\lambda(y))$. According to Lemma 6, the set of minimizers of $(\mathcal{P}_\lambda(y))$ reads $\mathcal{M} \subseteq x + \text{Ker}(\Phi)$, which is a convex set. We can therefore rewrite $(\mathcal{P}_\lambda(y))$ as

$$\min_{z \in \mathcal{M}} J(z).$$

Invoking Lemma 7 with $C = \mathcal{M}$, x is thus the unique minimizer if

$$\forall \delta \in \text{Ker}(\Phi) \setminus \{0\}, \quad J'(x, \delta) > 0.$$

Using Lemma 8 and the fact that $\text{Ker}(\Phi)$ is a subspace, this is equivalent to

$$\forall \delta \in \text{Ker}(\Phi) \setminus \{0\}, \quad \langle e_x, \delta_T \rangle + \langle P_S(f_x), \delta_S \rangle < J_f(-\delta_S).$$

which is **(NSP^S)**. \square

Proof of Corollary 1. Using [20, Theorem V.2.2.3], we know that

$$\eta \in \text{ri}(\partial J(x)) \Leftrightarrow J'(x, \delta) > \langle \eta, \delta \rangle \quad \forall \delta \text{ such that } J'(x, \delta) + J'(x, -\delta) > 0.$$

Applying this with $\eta = \Phi^* \alpha \in \text{ri}(\partial J(x))$, and using Lemma 8, we obtain

$$\Phi^* \alpha \in \text{ri}(\partial J(x)) \Leftrightarrow J'(x, \delta) > \langle \alpha, \Phi \delta \rangle \quad \forall \delta \text{ such that } J_{f_x}(\delta) + J_{f_x}(-\delta) > 0.$$

Moreover, since J_{f_x} and $\text{Ker}(J_{f_x}) = T_x = T$ from Proposition 4(iii), and (\mathcal{C}_T) holds, we get

$$\begin{aligned} \Phi^* \alpha \in \text{ri}(\partial J(x)) &\Leftrightarrow J'(x, \delta) > \langle \alpha, \Phi \delta \rangle \quad \forall \delta \notin T \\ &\Rightarrow J'(x, \delta) > 0 \quad \forall \delta \in \text{Ker}(\Phi). \end{aligned}$$

We conclude using Theorem 2. □

Proof of Theorem 3.

- (i) Let the dual vector $\alpha = (y - \Phi x)/\lambda$, and $\eta = \Phi^* \alpha \in \partial J(x)$ by Theorem 1(i). We then observe that

$$\begin{aligned} \eta \in \{ \eta \in \mathbb{R}^N : J_{f_x}^\circ(\eta_S - P_S(f_x)) < 1 \} &\Leftrightarrow \eta_S - P_S(f_x) \in \text{ri}(\partial J(x) - \{f_x\}) \\ &\Leftrightarrow \eta \in \text{ri}(\partial J(x)). \end{aligned}$$

Thus, applying Corollary 1 with such a dual vector yields the assertion.

- (ii) The proof is similar to (i) except that we invoke Theorem 1(ii). □

C. Proofs of Section 4

Proof of Proposition 7.

- (i) First, we have

$$\partial H(x) = \partial J(x) + \partial G(x),$$

Let $S^J = \text{span}(\partial J(x) - \eta^J)$ and $S^G = \text{span}(\partial G(x) - \eta^G)$, for any pair $\eta^J \in \partial J(x)$ and $\eta^G \in \partial G(x)$. Choosing $\eta^H = \eta^J + \eta^G \in \partial H(x)$ we have

$$\begin{aligned} S^H &= \text{span}(\partial H(x) - \eta^H) \\ &= \text{span}((\partial J(x) - \eta^J) + (\partial G(x) - \eta^G)) \\ &= \text{span}(\text{span}(\partial J(x) - \eta^J) + \text{span}(\partial G(x) - \eta^G)) \\ &= \text{span}(S^J \cup S^G). \end{aligned}$$

As a consequence we have $T^H = (S^H)^\perp = T^J \cap T^G$.

(ii) Moreover, since $T^H \perp S^J \cup S^G$ we have from Proposition 1(iii) that

$$\begin{aligned} e_H = P_{T^H}(\partial H(x)) &= P_{T^H}(\partial J(x) + \partial G(x)) \\ &= P_{T^H}(e_J + P_{S^J} \partial J(x) + e_G + P_{S^G} \partial G(x)) \\ &= P_{T^H}(e_J + e_G). \end{aligned}$$

(iii) As $f_x^J \in \text{ri } \partial J(x)$ and $f_x^G \in \text{ri } \partial G(x)$, it follows from [33, Corollary 6.6.2] that

$$f_x^H = f_x^J + f_x^G \in \text{ri } \partial J(x) + \text{ri } \partial G(x) = \text{ri } (\partial J(x) + \partial G(x)) = \text{ri } \partial H(x).$$

The antipromoting gauge associated to H is then

$$H_{f_x^H}^\circ = \gamma_{\partial H(x) - f_x^H} = \gamma_{(\partial J(x) - f_x^J) + (\partial G(x) - f_x^G)},$$

which is coercive and bounded on S^H according to Proposition 3. Invoking Lemma 3, we get the desired result since for any $\rho \geq 0$,

$$u \mapsto \rho J_{f_x^J}^\circ(u) + (1 - \rho) G_{f_x^G}^\circ(\eta - u) = \rho \gamma_{\partial J(x) - f_x^J}(u) + (1 - \rho) \gamma_{\partial G(x) - f_x^G}(\eta - u)$$

is bounded and continuous on $S^J \cap (S^G + \eta)$, for $\eta \in S^H = \text{span}(S^J + S^G)$ by (i). \square

Proof of Proposition 8. In the following, all operator bounds that appear are finite owing to the coercivity assumption on the involved gauges in Definition 9 of a PRG.

It is straightforward to see that the function $\Gamma^H = \max(\Gamma^J, \Gamma^G)$ is indeed a gauge, which is bounded and coercive on $T^H = T^J \cap T^G$. Moreover, given that both J and G are PRG at x with corresponding parameters ν_x^J and ν_x^G , we have with the advocated choice of Γ^H and ν_x^H ,

$$\Gamma^J(x - x') \leq \nu_x^J \quad \text{and} \quad \Gamma^G(x - x') \leq \nu_x^G,$$

for every $\forall x' \in T_x^H$ such that $\Gamma^H(x - x') \leq \nu_x^H$. It follows that:

- Since J and G are both PRG, then we have $T_x^J = T_{x'}^J$ and $T_x^G = T_{x'}^G$, and thus by Proposition 7(i)

$$T_x^H = T_x^J \cap T_x^G = T_{x'}^J \cap T_{x'}^G = T_{x'}^H = T^H.$$

- μ_x^H -**stability**: we have from Proposition 7(ii)

$$\begin{aligned} \Gamma^H(e_x^H - e_{x'}^H) &= \Gamma^H(P_{T^H}(e_x^J + e_x^G - e_{x'}^J - e_{x'}^G)) \\ &\leq \Gamma^H(P_{T^H}(e_x^J - e_{x'}^J)) + \Gamma^H(P_{T^H}(e_x^G - e_{x'}^G)) \\ &\leq \mathcal{M}_{\Gamma^J, \Gamma^H}(P_{T^H}) \Gamma^J(e_x^J - e_{x'}^J) + \mathcal{M}_{\Gamma^G, \Gamma^H}(P_{T^H}) \Gamma^G(e_x^G - e_{x'}^G) \\ &\leq (\mu_x^J \mathcal{M}_{\Gamma^J, \Gamma^H}(P_{T^H}) + \mu_x^G \mathcal{M}_{\Gamma^G, \Gamma^H}(P_{T^H})) \Gamma^H(x - x'), \end{aligned}$$

where we used μ_x^J - and μ_x^G -stability of J and G in the last inequality.

- τ_x^H -**stability**: the fact that $S^J \subseteq S^H$ and $S^G \subseteq S^H$ and subadditivity of gauges lead to

$$\begin{aligned}
& H_{f_x^H}^\circ (\mathbb{P}_{S^H}(f_x^H - f_{x'}^H)) \\
&= H_{f_x^H}^\circ (\mathbb{P}_{S^J}(f_x^J - f_{x'}^J) + \mathbb{P}_{S^G}(f_x^G - f_{x'}^G) + \mathbb{P}_{S^H}(e_x^J - e_{x'}^J) + \mathbb{P}_{S^H}(e_x^G - e_{x'}^G)) \\
&\leq H_{f_x^H}^\circ (\mathbb{P}_{S^J}(f_x^J - f_{x'}^J)) + H_{f_x^H}^\circ (\mathbb{P}_{S^G}(f_x^G - f_{x'}^G)) \\
&\quad + H_{f_x^H}^\circ (\mathbb{P}_{S^H}(e_x^J - e_{x'}^J)) + H_{f_x^H}^\circ (\mathbb{P}_{S^H}(e_x^G - e_{x'}^G)) . \tag{9}
\end{aligned}$$

According to Proposition 7(iii), we have

$$H_{f_x^H}^\circ (\mathbb{P}_{S^J}(f_x^J - f_{x'}^J)) = \inf_{\eta_1 + \eta_2 = \mathbb{P}_{S^J}(f_x^J - f_{x'}^J)} \max(J_{f_x^J}^\circ(\eta_1), G_{f_x^G}^\circ(\eta_2)) .$$

Since $\text{dom } J_{f_x^J}^\circ = S^J$, $(\eta_1, \eta_2) = (\mathbb{P}_{S^J}(f_x^J - f_{x'}^J), 0)$ is a feasible point of the last problem, and we get

$$H_{f_x^H}^\circ (\mathbb{P}_{S^J}(f_x^J - f_{x'}^J)) \leq J_{f_x^J}^\circ (\mathbb{P}_{S^J}(f_x^J - f_{x'}^J)) .$$

Moreover, as $e_x^J, e_{x'}^J \in T^J$ (see Proposition 1(ii)) and $S^J \subseteq S^H$, we have

$$\begin{aligned}
& \min_{\eta_1 \in T^J, \eta_2 \in S^J, \eta_1 + \eta_2 \in S^H} \|\eta_1 + \eta_2 - (e_x^J - e_{x'}^J)\|^2 \\
&= \min_{\eta_1 \in T^J, \eta_2 \in S^J, \eta_1 + \eta_2 \in S^H} \|\eta_1 - (e_x^J - e_{x'}^J)\|^2 + \|\eta_2\|^2 \\
&= \min_{\eta_1 \in T^J, \eta_2 \in S^J, \eta_1 \in S^H} \|\eta_1 - (e_x^J - e_{x'}^J)\|^2 + \|\eta_2\|^2 \\
&= \min_{\eta_1 \in S^H \cap T^J} \|\eta_1 - (e_x^J - e_{x'}^J)\|^2 .
\end{aligned}$$

That is

$$\mathbb{P}_{S^H}(e_x^J - e_{x'}^J) = \mathbb{P}_{S^H \cap T^J}(e_x^J - e_{x'}^J) .$$

Thus

$$H_{f_x^H}^\circ (\mathbb{P}_{S^H}(e_x^J - e_{x'}^J)) \leq \mathcal{M}_{\Gamma^J, H_{f_x^H}^\circ} (\mathbb{P}_{S^H \cap T^J}) \Gamma^J (e_x^J - e_{x'}^J) .$$

Similar reasoning leads to the following bounds

$$\begin{aligned}
& H_{f_x^H}^\circ (\mathbb{P}_{S^G}(f_x^G - f_{x'}^G)) \leq G_{f_x^G}^\circ (\mathbb{P}_{S^G}(f_x^G - f_{x'}^G)) , \\
& H_{f_x^H}^\circ (\mathbb{P}_{S^H}(e_x^G - e_{x'}^G)) \leq \mathcal{M}_{\Gamma^G, H_{f_x^H}^\circ} (\mathbb{P}_{S^H \cap T^G}) \Gamma^G (e_x^G - e_{x'}^G) .
\end{aligned}$$

Having this, we can continue to bound (9) as

$$\begin{aligned}
& H_{f_x^H}^\circ (\mathbb{P}_{S^H}(f_x^H - f_{x'}^H)) \\
& \leq J_{f_x^J}^\circ (\mathbb{P}_{S^J}(f_x^J - f_{x'}^J)) + G_{f_x^G}^\circ (\mathbb{P}_{S^G}(f_x^G - f_{x'}^G)) \\
& \quad + \mathcal{M}_{\Gamma^J, H_{f_x^H}^\circ} (\mathbb{P}_{S^H \cap T^J}) \Gamma^J (e_x^J - e_{x'}^J) + \mathcal{M}_{\Gamma^J, H_{f_x^H}^\circ} (\mathbb{P}_{S^H \cap T^G}) \Gamma^G (e_x^G - e_{x'}^G) \\
& \leq \tau_x^J \Gamma^J(x - x') + \tau_x^G \Gamma^G(x - x') + \mu_x^J \mathcal{M}_{\Gamma^J, H_{f_x^H}^\circ} (\mathbb{P}_{S^H \cap T^J}) \Gamma^J(x - x') \\
& \quad + \mu_x^G \mathcal{M}_{\Gamma^G, H_{f_x^H}^\circ} (\mathbb{P}_{S^H \cap T^G}) \Gamma^G(x - x') \\
& \leq \left(\tau_x^J + \tau_x^G + \mu_x^J \mathcal{M}_{\Gamma^J, H_{f_x^H}^\circ} (\mathbb{P}_{S^H \cap T^J}) + \mu_x^G \mathcal{M}_{\Gamma^G, H_{f_x^H}^\circ} (\mathbb{P}_{S^H \cap T^G}) \right) \Gamma^H(x - x'),
\end{aligned}$$

where the last two inequalities J and G follow from μ_x^J -, τ_x^J -, μ_x^G - and τ_x^G - stability of J and G .

- **ξ_x^H -stability:** Proposition 7(iii) again yields that for any $\eta \in S^H$

$$\begin{aligned}
H_{f_{x'}^H}^\circ(\eta) &= \inf_{\eta_1 + \eta_2 = \eta} \max(J_{f_{x'}^J}^\circ(\eta_1), G_{f_{x'}^G}^\circ(\eta_2)) \\
&\leq \max(J_{f_{x'}^J}^\circ(\bar{\eta}_1), G_{f_{x'}^G}^\circ(\bar{\eta}_2))
\end{aligned}$$

for any feasible $(\bar{\eta}_1, \bar{\eta}_2) \in S^J \times S^G \cap \{(\eta_1, \eta_2) : \eta_1 + \eta_2 = \eta\}$. Now both J and G are PRG, hence respectively ξ_x^J - and ξ_x^G -stable. Therefore, with the form of Γ^H we have

$$\begin{aligned}
J_{f_{x'}^J}^\circ(\bar{\eta}_1) &\leq (1 + \xi_x^J \Gamma^J(x - x')) J_{f_{x'}^J}^\circ(\bar{\eta}_1) \leq \beta J_{f_{x'}^J}^\circ(\bar{\eta}_1) \\
G_{f_{x'}^G}^\circ(\bar{\eta}_2) &\leq (1 + \xi_x^G \Gamma^G(x - x')) G_{f_{x'}^G}^\circ(\bar{\eta}_2) \leq \beta G_{f_{x'}^G}^\circ(\bar{\eta}_2),
\end{aligned}$$

where $\beta = 1 + \max(\xi_x^J, \xi_x^G) \Gamma^H(x - x')$. Whence we get

$$\max(J_{f_{x'}^J}^\circ(\eta_1), G_{f_{x'}^G}^\circ(\eta_2)) \leq \beta \max(J_{f_{x'}^J}^\circ(\bar{\eta}_1), G_{f_{x'}^G}^\circ(\bar{\eta}_2)).$$

Taking in particular

$$(\bar{\eta}_1, \bar{\eta}_2) \in \underset{\eta_1 + \eta_2 = \eta}{\text{Argmin}} \max(J_{f_{x'}^J}^\circ(\eta_1), G_{f_{x'}^G}^\circ(\eta_2))$$

we arrive at

$$H_{f_{x'}^H}^\circ(\eta) \leq \beta \inf_{\eta_1 + \eta_2 = \eta} \max(J_{f_{x'}^J}^\circ(\eta_1), G_{f_{x'}^G}^\circ(\eta_2)) = \beta H_{f_{x'}^H}^\circ(\eta).$$

This completes the proof. □

Proof of Proposition 9.

(i) One has $\partial J = D \circ \partial J_0 \circ D^*$, hence $S = DS_0 = \text{Im}(D_{S_0})$ and $T = S^\perp = \text{Ker}(D_{S_0}^*)$.

(ii) As $S = D\bar{S}_0 = De_0 + S$, we get from Proposition 1

$$\begin{aligned} e \in \underset{z \in \bar{S}}{\text{argmin}} \|z\| &= \underset{z - De_0 \in S}{\text{argmin}} \|z\| = De_0 + \underset{h \in S}{\text{argmin}} \|h + De_0\| \\ &= De_0 + P_S(-De_0) = (\text{Id} - P_S)De_0 = P_T De_0 . \end{aligned}$$

(iii) With such a choice of f_x , we have

$$\begin{aligned} f_{0,D^*x} \in \text{ri } \partial J_0(D^*x) &\Rightarrow Df_{0,D^*x} \in D \text{ri } \partial J_0(D^*x) \\ &\iff f_x \in \text{ri } D\partial J_0(D^*x) \iff f_x \in \text{ri } \partial J(x) . \end{aligned}$$

We follow the same lines as in the proof of Lemma 4, where we additionally invoke Proposition 4(ii) to get

$$\begin{aligned} J_{f_x}(d) &= \sigma_{\partial J(x) - f_x}(d) \\ &= \sigma_{D(\partial J_0(D^*x) - f_{0,D^*x})}(d) \\ &= \sigma_{\partial J_0(D^*x) - f_{0,D^*x}}(D^*d) \\ &= J_{0,f_{0,D^*x}}(D^*d) \\ &= J_{0,f_{0,D^*x}}(D_{S_0}^*d) . \end{aligned}$$

Note that J_{f_x} is indeed constant along affine subspaces parallel to $\text{Ker}(D_{S_0}^*) = S^\perp = T$. We now get that for every $\eta \in S = \text{Ker}(D_{S_0}^+)^\perp$

$$\begin{aligned} J_{f_x}^\circ(\eta) &= \sigma_{J_{f_x}(d) \leq 1}(\eta) \\ &= \sigma_{J_{0,f_{0,D^*x}}(D_{S_0}^*d) \leq 1}(\eta) \\ &= \left(\iota_{J_{0,f_{0,D^*x}}(w) \leq 1} \circ D_{S_0}^* \right)^*(\eta) \\ &= \inf_v \sigma_{J_{0,f_{0,D^*x}}(w) \leq 1}(v) \quad \text{s.t.} \quad D_{S_0}v = \eta \\ &= \inf_{z \in \text{Ker}(D_{S_0})} J_{0,f_{0,D^*x}}^\circ(D_{S_0}^+\eta + z) . \end{aligned}$$

The infimum is bounded and is attained necessarily at some $z \in \text{Ker}(D_{S_0}) \cap S_0 \neq \emptyset$ since $\text{dom } J_{0,f_{0,D^*x}}^\circ = S_0$ and $\text{Im}(D_{S_0}^+) = \text{Im}(D_{S_0}^*) \subset S_0$. Moreover, $\text{Ker}(D_{S_0}) \cap S_0 = \text{Ker}(D) \cap S_0$.

□

Proof of Proposition 10. In the following, all operator bounds that appear are finite owing to the coercivity assumption on the involved gauges in Definition 9 of a PRG.

- Let x' such that

$$\Gamma(x - x') \leq \frac{1}{\mathcal{M}_{\Gamma, \Gamma_0}(D^*)} \nu_{0,D^*x} .$$

Hence,

$$\Gamma_0(D^*x - D^*x') \leq \mathcal{M}_{\Gamma, \Gamma_0}(D^*)\Gamma(x - x') \leq \nu_{0, D^*x}$$

As J_0 is a PRG at D^*x , we have $T_{0, D^*x} = T_{0, D^*x'} = T_0$ and consequently, using Proposition 9(i), $T_x = \text{Ker}(D_{S_0, D^*x}^*) = \text{Ker}(D_{S_0, D^*x'}^*) = T_{x'} = T = S^\perp$.

- **μ_x -stability:** we now have

$$\begin{aligned} \Gamma(e_x - e'_x) &= \Gamma(\text{P}_T D(e_{0, D^*x} - e_{0, D^*x'})) && \text{Proposition 9(ii)} \\ &\leq \mathcal{M}_{\Gamma_0, \Gamma}(\text{P}_T D)\Gamma_0(e_{0, D^*x} - e_{0, D^*x'}) \\ &\leq \mu_{0, D^*x}\mathcal{M}_{\Gamma_0, \Gamma}(\text{P}_T D)\Gamma_0(D^*x - D^*x') && \text{using } \mu_{0, D^*x}\text{-stability of } J_0 \\ &\leq \mu_{0, D^*x}\mathcal{M}_{\Gamma_0, \Gamma}(\text{P}_T D)\mathcal{M}_{\Gamma, \Gamma_0}(D^*)\Gamma(x - x'). \end{aligned}$$

- **τ_x -stability:** since $f_{0, D^*x} \in \partial J_0(D^*x)$ and $f_{0, D^*x'} \in \partial J_0(D^*x')$, one has

$$f_{0, D^*x} - f_{0, D^*x'} = \text{P}_{S_0}(f_{0, D^*x} - f_{0, D^*x'}) + e_{0, D^*x} - e_{0, D^*x'}.$$

Thus, subadditivity yields

$$\begin{aligned} J_{f_x}^\circ(\text{P}_S(f_x - f_{x'})) &= J_{f_x}^\circ(\text{P}_S D(f_{0, D^*x} - f_{0, D^*x'})) \\ &\leq J_{f_x}^\circ(\text{P}_S D \text{P}_{S_0}(f_{0, D^*x} - f_{0, D^*x'})) + J_{f_x}^\circ(\text{P}_S D(e_{0, D^*x} - e_{0, D^*x'})). \end{aligned}$$

Using Proposition 9(iii) and τ_{0, D^*x} -stability of J_0 , we get the following bound on the first term

$$\begin{aligned} &J_{f_x}^\circ(\text{P}_S D \text{P}_{S_0}(f_{0, D^*x} - f_{0, D^*x'})) \\ &= \inf_{z \in \text{Ker}(D) \cap S_0} J_{0, f_{0, D^*x}}^\circ(D_{S_0}^+ \text{P}_S D \text{P}_{S_0}(f_{0, D^*x} - f_{0, D^*x'}) + z) \\ &\leq J_{0, f_{0, D^*x}}^\circ(D_{S_0}^+ \text{P}_S D \text{P}_{S_0}(f_{0, D^*x} - f_{0, D^*x'})) \\ &\leq \mathcal{M}_{J_{0, f_{0, D^*x}}^\circ, J_{0, f_{0, D^*x}}^\circ} (D_{S_0}^+ \text{P}_S D) J_{0, f_{0, D^*x}}^\circ(\text{P}_{S_0}(f_{0, D^*x} - f_{0, D^*x'})) \\ &\leq \tau_{0, D^*x} \mathcal{M}_{J_{0, f_{0, D^*x}}^\circ, J_{0, f_{0, D^*x}}^\circ} (D_{S_0}^+ \text{P}_S D) \Gamma_0(D^*x - D^*x') \\ &\leq \tau_{0, D^*x} \mathcal{M}_{J_{0, f_{0, D^*x}}^\circ, J_{0, f_{0, D^*x}}^\circ} (D_{S_0}^+ \text{P}_S D) \mathcal{M}_{\Gamma, \Gamma_0}(D^*)\Gamma(x - x'). \end{aligned}$$

Now, combining Proposition 9(iii) and μ_{0, D^*x} -stability of J_0 , we obtain the following bound on the second term

$$\begin{aligned} J_{f_x}^\circ(\text{P}_S D(e_{0, D^*x} - e_{0, D^*x'})) &\leq J_{0, f_{0, D^*x}}^\circ(D_{S_0}^+ \text{P}_S D(e_{0, D^*x} - e_{0, D^*x'})) \\ &\leq \mathcal{M}_{\Gamma_0, J_{0, f_{0, D^*x}}^\circ} (D_{S_0}^+ \text{P}_S D) \Gamma_0(e_{0, D^*x} - e_{0, D^*x'}) \\ &\leq \mu_{0, D^*x} \mathcal{M}_{\Gamma_0, J_{0, f_{0, D^*x}}^\circ} (D_{S_0}^+ \text{P}_S D) \mathcal{M}_{\Gamma, \Gamma_0}(D^*)\Gamma(x - x'). \end{aligned}$$

Combining these inequalities, we arrive at

$$\begin{aligned} J_{f_x}^\circ(\text{P}_S(f_x - f_{x'})) &\leq \left(\tau_{0, D^*x} \mathcal{M}_{J_{0, f_{0, D^*x}}^\circ, J_{0, f_{0, D^*x}}^\circ} (D_{S_0}^+ \text{P}_S D) \right. \\ &\quad \left. + \mu_{0, D^*x} \mathcal{M}_{\Gamma_0, J_{0, f_{0, D^*x}}^\circ} (D_{S_0}^+ \text{P}_S D) \right) \mathcal{M}_{\Gamma, \Gamma_0}(D^*)\Gamma(x - x'), \end{aligned}$$

whence we get τ_x -stability.

- ξ_x -**stability**: from Proposition 9(iii), we can write for any $\eta \in S$

$$\begin{aligned} J_{f_{x'}}^\circ(\eta) &= \inf_{z \in \text{Ker}(D) \cap S_0} J_{f_0, D^* x'}^\circ(D_{S_0}^+ \eta + z) \\ &\leq J_{f_0, D^* x'}^\circ(D_{S_0}^+ \eta + \bar{z}) \end{aligned}$$

for any $\bar{z} \in \text{Ker}(D) \cap S_0$.

Owing to $\xi_{0, D^* x}$ -stability of J_0 , and since $D_{S_0}^+ \eta \in S_0$, we have for any feasible $\bar{z} \in \text{Ker}(D) \cap S_0$

$$J_{f_0, D^* x'}^\circ(D_{S_0}^+ \eta + \bar{z}) \leq (1 + \xi_{0, D^* x} \Gamma_0(D^* x - D^* x')) J_{0, f_0, D^* x}^\circ(D_{S_0}^+ \eta + \bar{z}).$$

Taking in particular

$$\bar{z} \in \underset{z \in \text{Ker}(D) \cap S_0}{\text{Arginf}} J_{0, f_0, D^* x}^\circ(D_{S_0}^+ \eta + z)$$

we get the bound

$$\begin{aligned} J_{f_{x'}}^\circ(\eta) &\leq (1 + \xi_{0, D^* x} \Gamma_0(D^* x - D^* x')) \inf_{z \in \text{Ker}(D) \cap S_0} J_{0, f_0, D^* x}^\circ(D_{S_0}^+ \eta + z) \\ &= (1 + \xi_{0, D^* x} \Gamma_0(D^* x - D^* x')) J_{f_{x'}}^\circ(\eta) \\ &= (1 + \xi_{0, D^* x} \mathcal{M}_{\Gamma, \Gamma_0}(D^*) \Gamma(x - x')) J_{f_{x'}}^\circ(\eta), \end{aligned}$$

where we used again Proposition 9(iii) in the first equality. □

D. Proofs of Section 5

Proof of Theorem 4. This is a straightforward consequence of Theorem 3(ii) by constructing an appropriate dual certificate from $\mathbf{IC}(x_0)$. Denote $e = e_{x_0}$, $f = f_{x_0}$ and $S = T^\perp$. Taking the dual vector $\alpha = \Phi_T^{+,*} e$, we have on the one hand

$$\Phi_T^* \Phi_T^{+,*} e = e$$

since $e \in \text{Im}(\Phi_T^*)$.

On the other hand,

$$J_f^\circ(\Phi_S^* \Phi_T^{+,*} e - P_S f) = \mathbf{IC}(x_0) < 1. \quad \square$$

Proof of Theorem 5. To lighten the notation, we let $\varepsilon = \|w\|$, $\nu = \nu_{x_0}$, $\mu = \mu_{x_0}$, $\tau = \tau_{x_0}$, $\xi = \xi_{x_0}$, $f = f_{x_0}$.

The strategy is to construct a vector which is the unique solution to

$$\min_{x \in T} \frac{1}{2} \|y - \Phi x\|^2 + \lambda J(x), \quad (\mathcal{P}_\lambda^T(y))$$

and then to show that it is actually the unique solution to $(\mathcal{P}_\lambda(y))$ under the assumptions of Theorem 5.

The following lemma gives a convenient implicit equation satisfied by the unique solution to $(\mathcal{P}_\lambda^T(y))$.

Lemma 9. *Let $x_0 \in \mathbb{R}^N$ and denote $T = T_{x_0}$. Assume that (\mathcal{C}_T) holds. Then $(\mathcal{P}_\lambda^T(y))$ has exactly one minimizer \hat{x} , and the latter satisfies*

$$\hat{x} = x_0 + \Phi_T^+ w - \lambda(\Phi_T^* \Phi_T)^{-1} \tilde{e} \quad \text{where } \tilde{e} \in P_T(\partial J(\hat{x})). \quad (10)$$

Proof. Assumption (\mathcal{C}_T) implies that the objective in $(\mathcal{P}_\lambda^T(y))$ is strongly convex on the feasible set T , whence uniqueness follows immediately. By a trivial change of variable, $(\mathcal{P}_\lambda^T(y))$ be also rewritten in the unconstrained form

$$\hat{x} = \operatorname{argmin}_{x \in \mathbb{R}^N} \frac{1}{2} \|y - \Phi_T x\|^2 + \lambda J(P_T x).$$

Thus, using Proposition 5(i), \hat{x} has to satisfy

$$\Phi_T^*(y - \Phi_T \hat{x}) + \lambda \tilde{e} = 0,$$

for any $\tilde{e} \in P_T(\partial J(\hat{x}))$. Owing to the invertibility of Φ on T , i.e. (\mathcal{C}_T) , we obtain (10). \square

We are now in position to prove Theorem 5. This is be achieved in three steps:

Step 1: We show that in fact $T_{\hat{x}} = T$.

Step 2: Then, we prove that \hat{x} is the unique solution of $(\mathcal{P}_\lambda(y))$ using Theorem 3.

Step 3: We finally exhibit an appropriate regime on λ and ε for the above two statements to hold.

Step 1: Subspace equality. By construction of \hat{x} in $(\mathcal{P}_\lambda^T(y))$, it is clear that $T_{\hat{x}} \subseteq T$. The key argument now is to use that J is PRG at x_0 , and to show that

$$\Gamma(x_0 - \hat{x}) \leq \nu, \quad (11)$$

which in turn will imply subspace equality, i.e. $T_{\hat{x}} = T$ (see Definition 9).

We have from (10) and subadditivity that

$$\begin{aligned} \Gamma(x_0 - \hat{x}) &\leq \Gamma(-\Phi_T^+ w) + \lambda \Gamma((\Phi_T^* \Phi_T)^{-1} \tilde{e}) \\ &\leq \mathcal{M}_{\Gamma, \Gamma}((\Phi_T^* \Phi_T)^{-1}) \{ \Gamma(-\Phi_T^+ w) + \lambda \Gamma(\tilde{e}) \} \\ &\leq \mathcal{M}_{\Gamma, \Gamma}((\Phi_T^* \Phi_T)^{-1}) \{ \mathcal{M}_{\ell^2, \Gamma}(\Phi_T^*) \varepsilon + \alpha_0 \lambda \}. \end{aligned} \quad (12)$$

where $\alpha_0 = \Gamma(\tilde{e})$. Consequently, to show that (11) is verified, it is sufficient to prove that

$$A\varepsilon + B\lambda \leq \nu, \quad (C_1)$$

where we set the positive constants

$$\begin{aligned} A &= \mathcal{M}_{\Gamma, \Gamma}((\Phi_T^* \Phi_T)^{-1}) \mathcal{M}_{\ell^2, \Gamma}(\Phi_T^*), \\ B &= \alpha_0 \mathcal{M}_{\Gamma, \Gamma}((\Phi_T^* \Phi_T)^{-1}). \end{aligned}$$

Suppose for now that (C₁) holds and consequently, $T_{\hat{x}} = T$. Then decomposability of J on T (Theorem 1) implies that

$$\hat{e} = P_{T_{\hat{x}}}(\partial J(\hat{x})) = P_T(\partial J(\hat{x})) = \tilde{e},$$

where we have denote $\hat{e} = e_{\hat{x}}$. Thus (10) yields the following implicit equation

$$\hat{x} = x_0 + \Phi_T^+ w - \lambda(\Phi_T^* \Phi_T)^{-1} \hat{e}. \quad (13)$$

Step 2: \hat{x} is the unique solution of $(\mathcal{P}_\lambda(y))$. Recall that under condition (C₁), J is decomposable at \hat{x} and x_0 with the same model subspace T . Moreover, (13) is nothing but condition (2) in Theorem 3 satisfied by \hat{x} . To deduce that \hat{x} is the unique solution of $(\mathcal{P}_\lambda(y))$, it remains to show that (3) holds i.e.,

$$J_f^\circ(\lambda^{-1} \Phi_S^*(y - \Phi \hat{x}) - \hat{f}_S) < 1. \quad (14)$$

where we use the shorthand notations $\hat{f} = f_{\hat{x}}$ and $\hat{f}_S = P_S \hat{f}$.

Under condition (C₁), the ξ -stability property (6) of J at x_0 yields

$$J_f^\circ(\lambda^{-1} \Phi_S^*(y - \Phi \hat{x}) - \hat{f}_S) \leq (1 + \xi \Gamma(x_0 - \hat{x})) J_f^\circ(\lambda^{-1} \Phi_S^*(y - \Phi \hat{x}) - \hat{f}_S). \quad (15)$$

Furthermore, from (13), we can derive

$$\lambda^{-1} \Phi_S^*(y - \Phi \hat{x}) - \hat{f}_S = \Phi_S^* \Phi_T^{+,*} \hat{e} + \lambda^{-1} \Phi_S^* Q_T w - \hat{f}_S, \quad (16)$$

where $Q_T = \text{Id} - \Phi_T \Phi_T^+ = P_{\text{Ker}(\Phi_T^*)}$. Inserting (16) in (15), we obtain

$$J_f^\circ(\lambda^{-1} \Phi_S^*(y - \Phi \hat{x}) - \hat{f}_S) \leq (1 + \xi \Gamma(x_0 - \hat{x})) J_f^\circ(\Phi_S^* \Phi_T^{+,*} \hat{e} + \lambda^{-1} \Phi_S^* Q_T w - \hat{f}_S).$$

Moreover, subadditivity yields

$$\begin{aligned} J_f^\circ(\Phi_S^* \Phi_T^{+,*} \hat{e} + \lambda^{-1} \Phi_S^* Q_T w - \hat{f}_S) &\leq J_f^\circ(\Phi_S^* \Phi_T^{+,*} e - f_S) + J_f^\circ(\Phi_S^* \Phi_T^{+,*} (\hat{e} - e)) \\ &\quad + J_f^\circ(P_S(f - \hat{f})) + J_f^\circ(\lambda^{-1} \Phi_S^* Q_T w). \end{aligned} \quad (17)$$

We now bound each term of (17). In the first term, one recognizes

$$J_f^\circ(\Phi_S^* \Phi_T^{+,*} e - f_S) \leq \mathbf{IC}(x_0). \quad (18)$$

Appealing to the μ -stability property, we get

$$\begin{aligned} J_f^\circ(\Phi_S^* \Phi_T^{+,*}(\hat{e} - e)) &\leq \mathcal{M}_{\Gamma, J_f^\circ}(-\Phi_S^* \Phi_T^{+,*})\Gamma(e - \hat{e}) \\ &\leq \mu \mathcal{M}_{\Gamma, J_f^\circ}(-\Phi_S^* \Phi_T^{+,*})\Gamma(x_0 - \hat{x}). \end{aligned} \quad (19)$$

From τ -stability, we have

$$J_f^\circ(f_S - \hat{f}_S) \leq \tau \Gamma(x_0 - \hat{x}). \quad (20)$$

Finally, we use a simple operator bound to get

$$J_f^\circ(\lambda^{-1} \Phi_S^* Q_T w) \leq \frac{1}{\lambda} \mathcal{M}_{\ell^2, J_f^\circ}(\Phi_S^* Q_T) \varepsilon. \quad (21)$$

Following the same steps as for the bound (12), except using $\tilde{e} = \hat{e}$ here, gives

$$\Gamma(x_0 - \hat{x}) \leq \mathcal{M}_{\Gamma, \Gamma}((\Phi_T^* \Phi_T)^{-1}) \{ \mathcal{M}_{\ell^2, \Gamma}(\Phi_T^*) \varepsilon + \lambda \Gamma(\hat{e}) \}. \quad (22)$$

Plugging inequalities (18)-(22) into (15) we get the upper-bound

$$\begin{aligned} &J_f^\circ(\Phi_S^* \Phi_T^{+,*} \hat{e} + \lambda^{-1} \Phi_S^* Q_T w - \hat{f}_S) \\ &\leq (1 + \xi \Gamma(x_0 - \hat{x})) \left(\mathbf{IC}(x_0) + \Gamma(x_0 - \hat{x}) (\mu \mathcal{M}_{\Gamma, J_f^\circ}(-\Phi_S^* \Phi_T^{+,*}) + \tau) \right. \\ &\quad \left. + \frac{1}{\lambda} \mathcal{M}_{\ell^2, J_f^\circ}(\Phi_S^* Q_T) \varepsilon \right) \\ &\leq (1 + \xi(c_1 \varepsilon + \lambda c_2)) \left(\mathbf{IC}(x_0) + (c_1 \varepsilon + \lambda c_2) \bar{\mu} + \frac{1}{\lambda} c_4 \varepsilon \right) < 1, \end{aligned}$$

where we have introduced

$$\bar{\mu} = \mu c_3 + \tau \quad \text{and} \quad \alpha_1 = \Gamma(\hat{e}) = \Gamma(\tilde{e}) = \alpha_0$$

and

$$\begin{aligned} c_1 &= A, & c_2 &= \alpha_1 \mathcal{M}_{\Gamma, \Gamma}((\Phi_T^* \Phi_T)^{-1}), \\ c_3 &= \mathcal{M}_{\Gamma, J_f^\circ}(-\Phi_S^* \Phi_T^{+,*}), & c_4 &= \mathcal{M}_{\ell^2, J_f^\circ}(\Phi_S^* Q_T). \end{aligned}$$

If is then sufficient that

$$(1 + \xi(c_1 \varepsilon + \lambda c_2)) \left(\mathbf{IC}(x_0) + (c_1 \varepsilon + \lambda c_2) \bar{\mu} + \frac{1}{\lambda} c_4 \varepsilon \right) < 1, \quad (23)$$

for (3) in Theorem 3 to be in force.

In particular, if

$$C \varepsilon \leq \lambda$$

holds for some constant $C > 0$ to be fixed later, then inequality (23) is true if

$$P(\lambda) = a \lambda^2 + b \lambda + c > 0 \quad \text{where} \quad \begin{cases} a = -\xi \bar{\mu} (c_1/C + c_2)^2 \\ b = -(c_1/C + c_2) (\xi \mathbf{IC}(x_0) + \xi c_4/C + \bar{\mu}) \\ c = 1 - \mathbf{IC}(x_0) - c_4/C \end{cases} \quad (24)$$

Let us set the value of C to

$$C = \frac{2c_4}{1 - \mathbf{IC}(x_0)},$$

which, for $0 \leq \mathbf{IC}(x_0) < 1$, it ensures that $c = \frac{1 - \mathbf{IC}(x_0)}{2}$ is bounded and positive, and thus, the polynomial P has a negative and a positive root λ_{\max} equal to

$$\begin{aligned} \lambda_{\max} &= \frac{b}{2a} \varphi\left(-4 \frac{ac}{b^2}\right), \quad \begin{cases} a = -\xi \bar{\mu} ((1 - \mathbf{IC}(x_0))c_1 / (2c_4) + c_2)^2 \\ b = -((1 - \mathbf{IC}(x_0))c_1 / (2c_4) + c_2) (\bar{\mu} + (1 + \mathbf{IC}(x_0))\xi/2) \\ c = (1 - \mathbf{IC}(x_0))/2 \end{cases} \\ &= \frac{\bar{\mu} + (1 + \mathbf{IC}(x_0))\xi/2}{\xi \bar{\mu} ((1 - \mathbf{IC}(x_0))c_1 / c_4 + 2c_2)} \varphi\left(\frac{2\xi(1 - \mathbf{IC}(x_0))\bar{\mu}}{(\bar{\mu} + (1 + \mathbf{IC}(x_0))\xi/2)^2}\right) \\ &\geq \frac{1 - \mathbf{IC}(x_0)}{\xi} H(\bar{\mu}/\xi), \end{aligned}$$

where

$$\varphi(\beta) = \sqrt{1 + \beta} - 1, \quad \text{and} \quad H(\beta) = \frac{\beta + 1/2}{\beta(c_1/c_4 + 2c_2)} \varphi\left(\frac{2\beta}{(\beta + 1)^2}\right).$$

To get the above lower-bound on λ_{\max} , we used that φ is increasing (in fact strictly) and concave on \mathbb{R}_+ with $\varphi(1) = 0$, and that $\mathbf{IC}(x_0) \in [0, 1[$. Consequently, we can conclude that the bounds

$$\frac{2c_4}{1 - \mathbf{IC}(x_0)} \varepsilon \leq \lambda \leq \frac{1 - \mathbf{IC}(x_0)}{\xi} H(\bar{\mu}/\xi) \quad (C_2)$$

imply condition (23), which in turn yields (14).

Step 3: (C₁) and (C₂) are in agreement. It remains now that show the compatibility of (C₁) and (C₂), i.e. to provide appropriate regimes of λ and ε such that both conditions hold simultaneously. We first observe that (C₁) and the left-hand-side of (C₂) both hold for λ fulfilling

$$\lambda \leq C_0 \nu \quad \text{where} \quad C_0 = \left(\frac{A}{2c_4} + B\right)^{-1} \leq \left(\frac{1 - \mathbf{IC}(x_0)}{2c_4} A + B\right)^{-1}.$$

This updates (C₂) to the following ultimate range on λ

$$\frac{2c_4}{1 - \mathbf{IC}(x_0)} \varepsilon \leq \lambda \leq \min\left(C_0 \nu, \frac{1 - \mathbf{IC}(x_0)}{\xi} H(\bar{\mu}/\xi)\right).$$

Now in order to have an admissible non-empty range for λ , the noise level ε must be upper-bounded as

$$\varepsilon \leq \frac{1 - \mathbf{IC}(x_0)}{2c_4} \min\left(C_0 \nu, \frac{1 - \mathbf{IC}(x_0)}{\xi} H(\bar{\mu}/\xi)\right).$$

Finally, the constants provided in the statement of the theorem (and subsequent discussion) are as follows

$$A_T = 2c_4, \quad B_T = C_0, \quad D_T = c_3, \quad \text{and} \quad E_T = c_1/c_4 + 2c_2,$$

which completes the proof. \square

E. Proofs of Section 6

Proof of Proposition 11. The subdifferential of $\|\cdot\|_1$ reads

$$\partial\|\cdot\|_1(x) = \{\eta \in \mathbb{R}^N : \eta_{(I)} = \text{sign}(x_{(I)}) \quad \text{and} \quad \|\eta_{(I^c)}\|_\infty \leq 1\}.$$

The expressions of S_x , T_x , e_x and f_x follow immediately. Since $e_x \in \text{ri} \partial\|\cdot\|_1(x)$ and $\|\cdot\|_1$ is separable, it follows from Definition 8 that the ℓ^1 -norm is a strong gauge. Therefore $J_{f_x}^\circ = J^\circ = \|\cdot\|_\infty$, and Proposition 6 specializes to the stated subdifferential.

Turning to piecewise regularity, let $x' \in T$, i.e. $I(x') \subseteq I(x)$, and assume that

$$\|x - x'\|_\infty \leq \nu_x = (1 - \delta) \min_{i \in I} |x_i|, \quad \delta \in]0, 1[.$$

This implies that $\forall i \in I(x)$, $|x'_i| > \nu_x - \|x - x'\|_\infty \geq 0$, which in turn yields $I(x') = I(x)$, and thus $T_{x'} = T_x$. Since the sign is also locally constant on the restriction to T of the ℓ^∞ -ball centered at x of radius ν_x , one can choose $\mu_x = 0$. Finally $\tau_x = \xi_x = 0$ because $f_x = e_x$. \square

Proof of Proposition 13. The proof of the first part was given Section 2.2 and Section 2.3 where the ℓ^∞ -norm example was considered.

It remains to show piecewise regularity. Let $x' \in T$, and assume that

$$\|x - x'\|_1 \leq \nu_x = (1 - \delta) (\|x\|_\infty - \max_{j \notin I} |x_j|), \quad \delta \in]0, 1[.$$

This means that x' lies in the relative interior of the ℓ^1 -ball (relatively to T) centered at x of radius $\|x\|_\infty - \max_{j \notin I} |x_j|$. Within this ball, the support and the sign pattern restricted to the support are locally constant, i.e. $I(x) = I(x')$ and $\text{sign}(x_{(I(x))}) = \text{sign}(x'_{(I(x'))})$. Thus $T_{x'} = T_x = T$ and $e_{x'} = e_x$, and from the latter we deduce that $\mu_x = 0$. As $f_x = e_x$ we also conclude that $\tau_x = \xi_x = 0$, which completes the proof. \square

Proof of Proposition 14. Again, the proof of the first part was given Section 2.2 and Section 2.4 where the $\ell^1 - \ell^2$ -norm example was handled.

Let $x' \in T$, i.e. $I(x') \subseteq I(x)$, and $\nu_x = (1 - \delta) \min_{b \in I} \|x_b\|$, $\delta \in]0, 1[$. First, observe that the condition

$$\|x - x'\|_{\infty,2} = \max_{b \in \mathcal{B}} \|x_b - x'_b\| \leq \nu_x$$

ensures that for all $b \in I$

$$\|x'_b\| \geq \|x_b\| - \|x_b - x'_b\| > \nu_x - \|x - x'\|_{\infty,2} \geq 0,$$

and thus $I(x') = I(x)$, i.e. $T_{x'} = T_x$. Moreover, since the gauge is strong, one has $\tau_x = \xi_x = 0$. To establish the μ_x -stability we use the following lemma.

Lemma 10. *Given any pair of non-zero vectors u and v where, $\|u - v\| \leq \rho\|u\|$, for $0 < \rho < 1$, we have*

$$\left\| \frac{u}{\|u\|} - \frac{v}{\|v\|} \right\| \leq C_\rho \frac{\|u - v\|}{\|u\|},$$

where $C_\rho = \frac{\sqrt{2}}{\rho} \sqrt{1 - \sqrt{1 - \rho^2}} \in]1, \sqrt{2}[$.

Proof. Let $d = v - u$ and $\beta = \frac{\langle u, d \rangle}{\|u\|\|d\|} \in [-1, 1]$. We then have the following identities

$$\left\| \frac{u}{\|u\|} - \frac{v}{\|v\|} \right\|^2 = 2 - 2 \frac{\langle u, v \rangle}{\|u\|\|v\|} = 2 - 2 \frac{\|u\|^2 + \|u\|\|d\|\beta}{\|u\|\sqrt{\|u\|^2 + \|d\|^2 + 2\|u\|\|d\|\beta}}, \quad (25)$$

for non-zero vectors u and d , the unique maximizer of (25) is $\beta^* = -\|d\|/\|u\|$. Note that the assumption $\|d\|/\|u\| \leq \rho < 1$ assures β^* to comply with the admissible range of β and further, the argument of the square root will be always positive. Now, inserting β^* in (25), using concavity of $\sqrt{\cdot}$ on \mathbb{R}_+ , and that $\|d\|/\|u\| \leq \rho$, we can deduce the following bound

$$\begin{aligned} \left\| \frac{u}{\|u\|} - \frac{v}{\|v\|} \right\|^2 &\leq 2 - 2\sqrt{1 - \frac{\|d\|^2}{\|u\|^2}} = 2 - 2\sqrt{\left(1 - \frac{\|d\|^2}{\rho^2\|u\|^2}\right) + \frac{\|d\|^2}{\rho^2\|u\|^2}(1 - \rho^2)} \\ &\leq 2 - 2\left(\left(1 - \frac{\|d\|^2}{\rho^2\|u\|^2}\right) + \frac{\|d\|^2}{\rho^2\|u\|^2}\sqrt{1 - \rho^2}\right) \\ &= 2 - 2\left(1 - \frac{1 - \sqrt{1 - \rho^2}}{\rho^2} \frac{\|d\|^2}{\|u\|^2}\right) \\ &= 2 \frac{1 - \sqrt{1 - \rho^2}}{\rho^2} \frac{\|d\|^2}{\|u\|^2}. \end{aligned}$$

□

By definition of ν_x , we have $(1 - \delta)\|x_b\| > \nu_x$, for $\delta \in]0, 1]$, $\forall b \in I$, and thus $\|x_b - x'_b\| \leq \nu_x \leq (1 - \delta)\|x_b\|$. Lemma 10 then applies, and it follows that, $\forall b \in I$

$$\|\mathcal{N}(x_b) - \mathcal{N}(x'_b)\| \leq C_\rho \frac{\|x'_b - x_b\|}{\|x_b\|} \leq C_\rho \frac{\|x'_b - x_b\|}{\nu_x},$$

and therefore we get

$$\|\mathcal{N}(x) - \mathcal{N}(x')\|_{\infty,2} \leq \frac{C_\rho}{\nu_x} \|x' - x\|_{\infty,2},$$

which implies μ_x -stability for $\mu_x = C_\rho/\nu_x$. □

Proof of Proposition 15. In general, the subdifferential of J_0 reads

$$\partial J_0(u) = \left\{ \sum_{i \in I} \rho_i s_i a^i : \rho \in \Sigma_I, s_i \in \begin{cases} \{1\} & \text{if } u_i > 0 \\ [0, 1] & \text{if } u_i = 0 \\ \{0\} & \text{if } u_i < 0 \end{cases} \right\},$$

where Σ_I is the canonical simplex in $\mathbb{R}^{|I|}$, and $I = \{i \in \{1, \dots, N_H\} : (x_i)_+ = J_0(x)\}$.

- If $u_i \leq 0, \forall i \in \{1, \dots, N_H\}$, the above expression becomes

$$\partial J_0(u) = \left\{ \sum_{i \in I_0} \rho_i s_i a^i : \rho \in \Sigma_{I_0}, s_i \in [0, 1] \right\},$$

where $I_0 = \{i \in \{1, \dots, N_H\} : u_i = J_0(u) = 0\}$. Equivalently, $\partial J_0(u)$ is the intersection of the unit ℓ_1 ball and the positive orthant on $\mathbb{R}^{|I_0|}$. The expressions of S_u, T_u and e_u then follow immediately. $\partial J_0(u)$ then contains $e_u = 0$, but not in its relative interior. Choosing any f_u as advocated, we have $f_u \in \text{ri } \partial J_0(u)$. To get the antipromoting gauge, we use some calculus rules on gauges and apply Lemma 3 to get

$$J_{f_u}^\circ(\eta_{(I_0)}) = \inf_{\tau \geq 0, \tau(f_u)_i \geq -\eta_i \forall i \in I_0} \max(\|\tau f_u + \eta\|_1, \tau),$$

where the extra-constraints on τ come from the fact that $\partial J_0(u)$ is in the positive orthant, and the ℓ^1 norm is the gauge of the unit ℓ^1 -ball. We then have

$$\begin{aligned} J_{f_u}^\circ(\eta_{(I_0)}) &= \inf_{\tau \geq 0, \mu\tau \geq \max_{i \in I_0} -\eta_i} \max\left(\tau \sum_{i \in I_0} (\mu a^i + \eta_i), \tau\right) \\ &= \inf_{\tau \geq \max_{i \in I_0} (-\eta_i)_+ / \mu} \max\left(\tau \mu |I_0| + \sum_{i \in I_0} \eta_i, \tau\right). \end{aligned}$$

- Assume now that $u_i > 0$ for at least one $i \in \{1, \dots, N_H\}$. In such a case, $J_0(u) = \|u\|_\infty$, and the subdifferential becomes

$$\partial J_0(u) = \Sigma_{I_+},$$

where $I_+ = \{i \in \{1, \dots, N_H\} : u_i = J_0(u) \text{ and } u_i > 0\}$. The forms of S_u, T_u, e_u, f_u and the antipromoting gauge can then be retrieved from those of the ℓ^∞ -norm with $s_{(I_+)} = 1$ and $s_{(I_+^c)} = 0$.

For piecewise regularity, the parameters are derived following the same lines as for the ℓ^∞ -norm. Let $u' \in T$, and assume that

$$\|u - u'\|_1 \leq \nu_u = (1 - \delta) \left(\max_{i \in I_+} u_i - \max_{j \notin I_+, u_j > 0} u_j \right),$$

for $\delta \in]0, 1]$. This means that x' lies in the relative interior of the ℓ^1 -ball (relatively to T) centered at x of radius

$$\max_{i \in I_+} u_i - \max_{j \notin I_+, u_j > 0} u_j = \|u\|_\infty - \max_{j \notin I_+, u_j > 0} |u_j|$$

Within this set, one can observe that the set I_+ associated to u is constant. Moreover, the sign pattern is also constant leading to the fact that $T_{u'} = T_u = T$. Hence, we deduce as in the ℓ^∞ -case that $\mu_u = \tau_u = \xi_u = 0$. \square

Proof of Proposition 16. The subdifferential of the nuclear norm is a classical result in convex analysis of spectral functions, see e.g. [42, 22]. The expressions of the subspaces T_x , S_x and e_x follow immediately. Since the nuclear norm is a strong gauge, we get from Proposition 6 that the antipromoting gauge is the spectral norm.

Let's turn to piecewise regularity. The proof is a straightforward adaptation of the arguments in [2, Proposition 16]. The latter result was stated for the spectral norm, but remains valid for the nuclear norm. \square

References

- [1] F. Bach. Consistency of the group lasso and multiple kernel learning. *The Journal of Machine Learning Research*, 9:1179–1225, 2008.
- [2] F. Bach. Consistency of trace norm minimization. *The Journal of Machine Learning Research*, 9:1019–1048, 2008.
- [3] F. Bach. Structured sparsity-inducing norms through submodular functions. *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2010.
- [4] S. Bakin. *Adaptive regression and model selection in data mining problems*. PhD thesis, Australian National University, 1999.
- [5] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411, 2004.
- [6] J. A. Cadzow. Algorithm for the minimum-effort problem. *IEEE Trans. Autom. Control.*, 16(1):60–63, Feb 1971.
- [7] E. Candès and B. Recht. Simple bounds for recovering low-complexity models. *Mathematical Programming*, pages 1–13, 2012.
- [8] E.J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational mathematics*, 9(6):717–772, 2009.
- [9] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky. The convex geometry of linear inverse problems. *Foundations of Computational mathematics*, 12(6):805–849, Dec. 2012.

- [10] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61, 1999.
- [11] C. Chesneau and M. Hebiri. Some theoretical results on the grouped variables lasso. *Mathematical Methods of Statistics*, 17(4):317–326, 2008.
- [12] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001.
- [13] D.L. Donoho and J. Tanner. Counting the faces of randomly-projected hypercubes and orthants, with applications. *Discrete & computational geometry*, 43(3):522–541, 2010.
- [14] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse problems*, 23(3):947, 2007.
- [15] M. J. Fadili, G. Peyré, S. Vaiter, C-A. Deledalle, and J. Salmon. Stable recovery with analysis decomposable priors. In *Proc. SampTA*, 2013.
- [16] M. Fazel. *Matrix Rank Minimization with Applications*. PhD thesis, Stanford University,, 2002.
- [17] J.-J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344, 2004.
- [18] M. Grasmair. Linear convergence rates for tikhonov regularization with positively homogeneous functionals. *Inverse Problems*, 27(7):075014, 2011.
- [19] E. Grave, G. Obozinski, and F. Bach. Trace lasso: a trace norm regularization for correlated designs. *arXiv preprint arXiv:1109.1990*, 2011.
- [20] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis And Minimization Algorithms*, volume I and II. Springer, 2001.
- [21] H. Jégou, T. Furon, and J.-J. Fuchs. Anti-sparse coding for approximate nearest neighbor search. In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 2029–2032. IEEE, 2012.
- [22] A.S. Lewis. The convex analysis of unitarily invariant matrix functions. *Journal of Convex Analysis*, 2(1/2):173–183, 1995.
- [23] H. Liu and J. Zhang. Estimation consistency of the group lasso and its applications. *Journal of Machine Learning Research*, 5:376–383, 2009.
- [24] Y. Lyubarskii and R. Vershynin. Uncertainty principles and vector quantization. *Information Theory, IEEE Transactions on*, 56(7):3491–3501, 2010.
- [25] O. Mangasarian and B. Recht. Probability of unique integer solution to a system of linear equations. *European Journal of Operational Research*, 214(1):27–30, 2011.

- [26] S. Nam, M.E. Davies, M. Elad, and R. Gribonval. The cosparsity analysis model and algorithms. *Applied and Computational Harmonic Analysis*, 34(1):30–56, 2013.
- [27] S. Negahban, P. Ravikumar, M.J. Wainwright, and B. Yu. A unified framework for high-dimensional analysis of m -estimators with decomposable regularizers. *arXiv preprint arXiv:1010.2731*, 2010.
- [28] G. Obozinski and F. Bach. Convex relaxation for combinatorial penalties. *arXiv preprint arXiv:1205.1240*, 2012.
- [29] S. Oymak, A. Jalali, M. Fazel, Y.C. Eldar, and B. Hassibi. Simultaneously structured models with application to sparse and low-rank matrices. *arXiv preprint arXiv:1212.3753*, 2012.
- [30] S. Petry and G. Tutz. Shrinkage and variable selection by polytopes. *Journal of Statistical Planning and Inference*, 142(1):48–64, 2012.
- [31] N. Rao, B. Recht, and R. Nowak. Signal recovery in unions of subspaces with applications to compressive imaging. *arXiv preprint arXiv:1209.3079*, 2012.
- [32] B. Recht, M. Fazel, and P.A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review*, 52(3):471–501, 2010.
- [33] R.T. Rockafellar. *Convex analysis*, volume 28. Princeton University Press, 1996.
- [34] V. Roth and B. Fischer. The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms. In *Proceedings of the International Conference on Machine Learning (ICML)*, volume 104, 2008.
- [35] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.
- [36] 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 Journal on Numerical Analysis*, 42(2):686–713, 2004.
- [37] C. Studer, W. Yin, and R.G. Baraniuk. Signal representations with minimum ℓ_∞ -norm. In *Proc. 50th Ann. Allerton Conf. on Comm. Control, and Computing (Allerton)*, 2012.
- [38] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1):267–288, 1996.
- [39] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2004.

- [40] S. Vaiter, G. Peyré, C. Dossal, and M. J. Fadili. Robust sparse analysis regularization. *IEEE Transactions on Information Theory*, 59(4):2001–2016, 2013.
- [41] S. Vaiter, G. Peyré, and M. J. Fadili. Robust polyhedral regularization. In *Proc. SampTA*, 2013.
- [42] G.A. Watson. Characterization of the subdifferential of some matrix norms. *Linear Algebra and its Applications*, 170:33–45, 1992.
- [43] F. Wei and J. Huang. Consistent group selection in high-dimensional linear regression. *Bernoulli*, 16(4):1369–1384, 2010.
- [44] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2005.