Model Selection with Low Complexity Priors

SAMUEL VAITER
CNRS, CEREMADE, Université Paris-Dauphine
Corresponding author: vaiter@ceremade.dauphine.fr

MOHAMMAD GOLBABAEE
CNRS, CEREMADE, Université Paris-Dauphine
golbabaee@ceremade.dauphine.fr

JALAL FADILI
GREYC, CNRS-ENSICAEN-Université de Caen
Jalal.Fadili@greyc.ensicaen.fr

AND

GABRIEL PEYRÉ
CNRS, CEREMADE, Université Paris-Dauphine
peyre@ceremade.dauphine.fr

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 partly smooth functions relative to a subspace. These are convex, non-negative, closed and finite-valued 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), as well as several others including the ℓ^∞ norm. We also show that the set of partly smooth functions relative to a subspace is closed under addition and pre-composition by a linear operator, which allows to cover mixed regularization, and the so-called analysis-type priors (e.g. total variation, fused Lasso, finite-valued 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, and on an analysis of the performance of ℓ^∞ regularization in a compressed sensing scenario.

Keywords: Convex regularization, Inverse problems, Model selection, Partial smoothness, Compressed Sensing, Sparsity, Total variation.

1. Introduction

1.1 *Regularization of Linear Inverse Problems*

Linear 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

$$y = \Phi x_0 + w, \tag{1.1}$$

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 (1.1) 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 (1.1) 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

$$x^\star \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}} \frac{1}{2}\|y - \Phi x\|^2 + \lambda J(x), \tag{$\mathscr{P}_\lambda(y)$}$$

where the first term expresses 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 ($\mathscr{P}_\lambda(y)$), 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 though we focus our attention on quadratic data fidelity for simplicity, our analysis carries over to more general fidelity terms of the form $F \circ \Phi$, for F smooth and strongly convex.

When there is no noise in the observations, i.e. $w = 0$ in (1.1), the equality-constrained minimization problem should be solved

$$x^\star \in \underset{x \in \mathbb{R}^N}{\operatorname{Argmin}}\, J(x) \quad \text{subject to} \quad \Phi x = y. \tag{$\mathscr{P}_0(y)$}$$

In this paper, we consider the general case where the function J is convex, non-negative and finite-valued[1], hence everywhere continuous. 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 [Tib96], the discrete total variation semi-norm [ROF92], the $\ell^1 - \ell^2$ norm to induce block/group sparsity [YL05], or finite polyhedral gauges [VPF13].

Assuming furthermore that J enjoys a partial smoothness property (to be defined in Section 5) relative to a model subspace associated to x_0, our goal in this paper is to provide a unified analysis of exact and robust recovery guarantees of that subspace by solving ($\mathscr{P}_\lambda(y)$) from the partial measurements in (1.1). As a by-product, this will also entail a control on the ℓ^2-recovery error.

1.2 *Contributions*

Our main contributions are as follows.

[1] Finite-valued means that $J(x) < +\infty$ for every $x \in \mathbb{R}^N$.

1.2.1 *Subdifferential Decomposability of Convex Functions.* Building upon Definition 3, which introduces the model subspace T_x at x, we provide an equivalent description of the subdifferential of a finite-valued convex function at x in Theorem 1. Such a description isolates and highlights a key property of a regularizer, namely *decomposability*. In turn, this property allows to rewrite the first-order minimality conditions of $(\mathscr{P}_\lambda(y))$ and $(\mathscr{P}_0(y))$ in a convenient and compact way, and this lays the foundations of our subsequent developments.

1.2.2 *Uniqueness.* In Theorem 2, we state a sharp sufficient condition, dubbed the Strong Null Space Property, to ensure that the solution of $(\mathscr{P}_\lambda(y))$ or $(\mathscr{P}_0(y))$ 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 $(\mathscr{P}_\lambda(y))$ and $(\mathscr{P}_0(y))$.

1.2.3 *Partly Smooth Functions Relative to a Subspace.* 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 *partly smooth function relative to a subspace*.

We show in particular that two important operations preserve partial smoothness relative to a subspace. In Proposition 9 and Proposition 11, we show that it is preserved under addition and pre-composition by a linear operator. Consequently, more intricate regularizers can be built starting from simple functions, e.g. ℓ^1-norm, which are known to be partly smooth relative to a subspace (see the review given in Section 7).

1.2.4 *Exact and Robust Subspace Recovery.* This is the core contribution of the paper. Assuming the function is partly smooth relative to a subspace, we show in Theorem 6 that under a generalization of the irrepresentability condition [Fuc04], 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 $(\mathscr{P}_\lambda(y))$ has a unique solution $x^\star$, which turns out to live in the same subspace as x_0. Clearly, solving $(\mathscr{P}_\lambda(y))$ for this regime of noise and λ allows to stably recover the subspace model underlying x_0. In turn, this yields a control on ℓ^2-recovery error within a factor of the noise level, i.e. $\|x^\star - x_0\| = O(\|w\|)$. In the noiseless case, the irrepresentability condition implies that x_0 is exactly identified by solving $(\mathscr{P}_0(y))$.

1.2.5 *Compressed Sensing with ℓ^∞ Norm Regularization.* To illustrate the usefulness of our findings, we apply this model recovery result to the case of the ℓ^∞ norm in Section 8. This regularization is known to promote anti-sparse (flat) vectors x_0. While there exists previous works on ℓ^2-stable recovery with ℓ^∞ regularization from random measurements, it is the first result to assess stable recovery of the anti-sparse model associated to x_0, which is an important additional information. Our result shows that stable model recovery operates at a different regime compared to ℓ^2-stable recovery in terms of bounds on the number of generic measurements as a function of the anti-sparsity level. This somehow contrasts with classical results in sparse recovery where it is known that both types of stable recovery hold at comparable bounds (up to logarithmic terms), see Section 1.4.4.

1.3 *Novelties and Limitations*

Before providing a detailed comparison with the state-of-the-art in the following section, we would like to stress why our contributions are not just unifying with an unprecedented level of generality, but they also allow to go beyond classical sparsity-type penalties and to tackle many regularizers that are not covered by the current literature.

First of all, it is important to note that our contributions on both subdifferential decomposability (Section 1.2.1) and uniqueness characterization (Section 1.2.2) are generic and do not put any constraint on the regularizer J (beside being convex and finite-valued). These results thus generalize many well-known ones that are scattered in the literature and derived for specific sparsity-enforcing priors (such as ℓ^1 or $\ell^1-\ell^2$ norms).

Our main contribution (Section (1.2.4)), which proves that the low-dimensional model subspace underlying x_0 can be robustly recovered from noisy measurements, is only valid for convex functions that are so-called partly-smooth at x_0 to a subspace. Loosely speaking, a partly smooth function behaves smoothly along a manifold, and transverse to it, they behave sharply. Partial smoothness offers a powerful framework in variational analysis to study sensitivity of optimization problems to perturbations of their parameters, and in particular, stability of the partial smoothness manifold. This is exactly our setting where the goal is to understand when the model manifold (hopefully low-dimensional) underlying the original object x_0 can be stably recovered from partial and noisy measurements. Thus partial smoothness of the regularizer appears a natural and wise assumption. In this paper, we focus on the case where the partial smoothness manifold is actually a subspace. While this may appear restrictive, it nevertheless allows us to provide a detailed analysis, where the constants in the stability bounds are made explicit. These results hold similarly for the case of affine manifolds. However, considering arbitrary (possibly curved) manifolds is more involved and not covered by our analysis here. Removing this assumption is possible (see for instance the recent work [VPF14] and the discussion in the following section), but the price to pay is that the stability bounds do not give access to explicit constants.

A typical novel application of our results is recovery of anti-sparse signals from partial random measurements using ℓ^∞ regularization, i.e. ℓ^∞ compressed sensing (see Section 1.2.5), which cannot be handled by existing previous works. This is however only the tip of the iceberg, and many more applications could be found. Typical other illustrative examples include polyhedral regularizations, and composition of the $\ell^1-\ell^2$ norm with a linear operator, as is the case for instance for the isotropic total variation which is very popular in image processing.

1.4 *Related Work*

1.4.1 *Decomposability.* In [CR12], the authors introduced a notion of decomposable norms. In fact, we show that their regularizers are a subclass of ours that corresponds to strong decomposability in the sense of the Definition 6, besides symmetry since norms are symmetric gauges. Moreover, their definition involves two conditions, the second of which turns out to be an intrinsic property implied by polarity rather than an assumption; see the discussion after Proposition 7. 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 [CR12] deals only with identifiability in the noiseless case. Their work was extended in [OJF$^+$12] when J is the sum of decomposable norms.

1.4.2 *Convergence rates.* In the inverse problems literature, convergence (stability) rates have been derived in [BO04] with respect to the Bregman divergence for general convex regularizations J. The author in [Gra11] 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 functions. In [FPV$^+$13], an ℓ^2-stability result for decomposable norms (in the sense of [CR12]) 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.

1.4.3 *Model selection.* There is large body of previous works on the problem of the model selection properties (sometimes referred to as model consistency) of low-complexity regularizers. These previous works are targeting specific regularizers, most notably sparsity, group sparsity and low rank. We thus refer to Section 7 for a discussion of these relevant previous works. A distinctive feature of our analysis is that it is generic, so it covers all these special cases, and many more. Note however that is does not cover the nuclear norm, because its associated manifolds are not linear (they are indeed composed of algebraic manifolds of low rank matrices). We have recently proposed an extension of our results to this more general non-linear case in [VPF14]. Note however that this new analysis uses a different proof technique, and is not able to provide explicit values for the constant involved in the robustness to noise.

1.4.4 *Compressed sensing.* Arguments based on the Gaussian width were used in [CRPW12] to provide sharp estimates of the number of generic measurements required for exact and ℓ^2-stable recovery of atomic set models from partial Gaussian measurements by solving a constrained form of $(\mathscr{P}_\lambda(y))$ regularized by an atomic norm. The atomic norm framework was then exploited in [RRN12] in the particular case of the group Lasso and union of subspace models. This was further generalized in [ALMT13] who developed for the noiseless case reliable predictions about the quantitative aspects of the phase transition in convex regularized linear inverse problems with Gaussian measurements. The location and width of the transition are controlled by the statistical dimension of the descent cone of the regularizer at the original vector x_0. When the noise is also Gaussian with a small enough variance, [OTH13] proposes a formula for calculating the normalized squared error for the estimator provided by solving $(\mathscr{P}_\lambda(y))$ with a general convex regularizer. All these works are however restricted to a random (compressed sensing) scenario.

A notion of decomposability closely related to that of [CR12], but different, was first proposed in [NRWY10]. 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 stronger assumptions than ours (typically a restricted strong convexity which becomes a type of restricted eigenvalue property for linear regression with quadratic data fidelity).

1.5 *Paper Organization*

The outline of the paper is the following. Section 2 provides a short recap on convex analysis. Section 3 fully characterizes the canonical decomposition of the subdifferential of a convex function with respect to the subspace model at x. Sufficient conditions ensuring uniqueness of the minimizers to $(\mathscr{P}_\lambda(y))$ and $(\mathscr{P}_0(y))$ are provided in Section 4. In Section 5, we introduce the notion of a partly smooth function relative to a subspace and show that this property is preserved under addition and pre-composition by a linear operator. Section 6 is dedicated to our main result, namely theoretical guarantees for exact subspace recovery in the presence of noise, and identifiability in the noiseless case. Section 7 exemplifies our results on several previously studied priors, and a detailed discussion on the relation with respect to relevant previous work is provided. Section 8 delivers a bound for the sampling complexity to guarantee exact recovery of the model subspace of antisparsity minimization from noisy Gaussian measurements. Some conclusions and possible perspectives of this work are drawn in Section 9. The proofs of our results are collected in the appendix.

2. A Short Tour of Convex Analysis

This sections aims to provide a short review of important tools from convex analysis that are used in this paper. A comprehensive account can be found in [Roc96, HUL01].

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

$$x_T = \mathrm{P}_T x \quad \text{and} \quad \Phi_T = \Phi \mathrm{P}_T \,.$$

For a subset I of $\{1,\ldots,N\}$, 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 $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\}$.

2.1 *Sets*

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

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

For instance, the affine hull of a segment in $\mathbb{R}^2$ is the straight line containing this segment. It is a translate of its *parallel subspace* $\mathrm{par}\, C$, i.e. $\mathrm{par}\, C = \mathrm{aff}\, C - x = \mathrm{span}(C - x)$ for any $x \in C$, where $\mathrm{span}\, C$ is the linear hull of C.

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

2.2 *Functions*

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

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

Let C be 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

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

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 lsc and sublinear. It is non-negative if $0 \in C$. Moreover, we have the following.

LEMMA 1 Let C be a non-empty set.

(i) σ_C is lsc and sublinear.

(ii) σ_C is finite-valued if and only if C is bounded.

(iii) If $0 \in C$, then σ_C is non-negative.

(iv) If C is convex and $0 \in C$, then σ_C is constant along all affine subspaces parallel to $\mathrm{par} C$.

(v) If C is convex and compact with $0 \in \mathrm{ri} C$, then σ_C is finite-valued, $\mathrm{Ker}\, \sigma_C = (\mathrm{par} C)^\perp$ and σ_C is coercive on $\mathrm{par} C$.

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

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

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, $\gamma_C(x) = +\infty$ in case of emptiness of the set over which the infimum is computed. γ_C is a non-negative, lsc and sublinear function. It is moreover finite everywhere, hence continuous, if, and only if, C has the origin as an interior point, see Lemma 2 for details.

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

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

An element of $\partial f(x)$ is a subgradient. If the convex function f is 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 $x \in \mathrm{dom} f$ 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 $\delta \mapsto f'(x, \cdot)$ exists and is sublinear. When f has also full domain, then for any $x \in \mathbb{R}^N$, $\partial f(x)$ 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: $x^\star$ is the global minimizer of a convex function f if, and only if, $0 \in \partial f(x)$.

2.3 *Gauges*

We start by collecting some important properties of gauges and their polars. A comprehensive account on them can be found in [Roc96].

Lemma 2, in particular item (ii), is a 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.

LEMMA 2

(i) γ_C is a non-negative, lsc and sublinear function.

(ii) C is the unique closed convex set containing the origin such that

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

(iii) γ_C is finite everywhere if, and only if, $0 \in \operatorname{int} C$, in which case γ_C is continuous.

(iv) $\operatorname{Ker} \gamma_C = \{0\}$ if, and only if, C is compact.

(v) γ_C is finite and coercive on $\operatorname{dom} \gamma_C = \operatorname{par} C$ if, and only if, C is compact and $0 \in \operatorname{ri} C$. In particular, γ_C is finite everywhere and coercive if, and only if, C is compact and $0 \in \operatorname{int} C$.

Observe that 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 1 (Polar set) Let C be a non-empty convex set. The set C° given by

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

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 2 (Polar Gauge) The polar of a gauge γ_C is the function γ_C° defined by

$$\gamma_C^\circ(u) = \inf\left\{\mu \geqslant 0 : \ \langle x, u\rangle \leqslant \mu\gamma_C(x), \forall x\right\} .$$

Observe that gauges polar to each other have the property

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

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

LEMMA 3 Let $C \subseteq \mathbb{R}^N$ be a closed convex set containing 0. Then,

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

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

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

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

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

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 5.

LEMMA 4 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 finite 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 5 Let C be a compact convex set containing 0, and D a linear operator. Then, for every $x \in \operatorname{Im}(D)$

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

When it is also assumed that $0 \in \operatorname{ri} C$, using Lemma 2(v), one can observe that the infimum is finite if $(D^+ x + \operatorname{Ker}(D)) \cap \operatorname{par} C \neq \emptyset$.

2.4 *Set-valued mappings*

We need in this paper some basic facts on set-valued mappings. A comprehensive account can be found in [AF09]. A *set valued-mapping* $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is characterized by its graph, i.e. by the subset of $X \times Y$ defined by

$$\operatorname{graph}(F) = \{(x,y) \in X \times Y : \; y \in F(x)\} \, .$$

The domain of F, $\operatorname{dom} F$, is the set of points $x \in \mathbb{R}^n$ such that $F(x) \neq \emptyset$.

A set-valued mapping F is *Lipschitz* relative to a non-empty set U in $\mathbb{R}^n$ if $U \subset \operatorname{dom} F$, F is closed-valued on U and there exists $\beta \geqslant 0$ such that

$$F(x) \subseteq F(x') + \beta \|z - z'\| \mathbb{B}(0), \quad \text{for all } x, x' \in U \; ,$$

where $\mathbb{B}(0)$ is the unit ball of $\mathbb{R}^m$.

We end by showing that Lipschitz continuity of F transfers to that of the associated gauge.

LEMMA 6 Let $F : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ be β-Lipschitz on a compact set U, and assume that for every point $x \in U$, $F(x)$ is a compact convex set containing the origin as a relative interior point. Then, for any x, x' in U, and $u \in \operatorname{par}(F(x)) \cap \operatorname{par}(F(x'))$, there exists a constant constant $C < +\infty$ such that the mapping $x \in U \mapsto \gamma_{F(x)}(u)$ is $C\beta\|u\|$-Lipschitz continuous.

2.5 *Operator norm*

Let J_1 and J_2 be two finite-valued gauges defined on two vector spaces V_1 and V_2, and $A : V_1 \to V_2$ a linear map. The *operator bound* $|\!|\!| A |\!|\!|_{J_1 \to J_2}$ of A between J_1 and J_2 is given by

$$|\!|\!| A |\!|\!|_{J_1 \to J_2} = \sup_{J_1(x) \leqslant 1} J_2(Ax).$$

Note that $|||A|||_{J_1 \to J_2} < +\infty$ if, and only if $A\,\mathrm{Ker}(J_1) \subseteq \mathrm{Ker}(J_2)$. In particular, if J_1 is coercive (i.e. $\mathrm{Ker}\,J_1 = \{0\}$ from Lemma 2(v)), then $|||A|||_{J_1 \to J_2}$ is finite. As a convention, $|||A|||_{J_1 \to \|\cdot\|_p}$ is denoted as $|||A|||_{J_1 \to \ell^p}$. An easy consequence of this definition is the fact that for every $x \in V_1$,

$$J_2(Ax) \leqslant |||A|||_{J_1 \to J_2} J_1(x).$$

3. Model Subspace and Decomposability

The purpose of this section is to introduce one of the main concepts used throughout this paper, namely the *model subspace* associated to a convex function. The main result, Theorem 1, proves that the subdifferential of any convex function exhibits a decomposability property with respect to this subspace.

In the case of ℓ^1-norm, the following result is well-known.

FACT 1 (Decomposability of ℓ^1) Let $x \in \mathbb{R}^N$. Then the subdifferential of $\|\cdot\|_1$ at x reads

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

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

In plain words, this result decomposes the subdifferential of the ℓ^1-norm at a point x into a single-valued part characterized by the sign vector of the active components of x, i.e. those indexed by its support I, and a set-valued part corresponding to the non-active components indexed by I^c. In the following section, we show how to generalize this splitting to any finite-valued convex function.

3.1 *Model Subspace Associated to a Convex Function*

Let J be our regularizer, i.e. a finite-valued convex function.

DEFINITION 3 (Model Subspace) For any vector $x \in \mathbb{R}^N$, denote $\bar{S}_x$ the affine hull of the subdifferential of J at x

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

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

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

Let

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

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

When J is differentiable at x, i.e. $\partial J(x) = \{\nabla J(x)\}$, $e_x = \nabla J(x)$ and $T_x = \mathbb{R}^N$. Note that the decomposition of $\mathbb{R}^N$ as a sum of the two orthogonal subspaces T_x and S_x is also the core idea underlying the $\mathscr{U} - \mathscr{V}$-decomposition/theory developed in [LOS00].

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) $e_x \in T_x \cap \bar{S}_x$.

(ii) $\bar{S}_x = \left\{ \eta \in \mathbb{R}^N : \ \eta_{T_x} = e_x \right\}$.

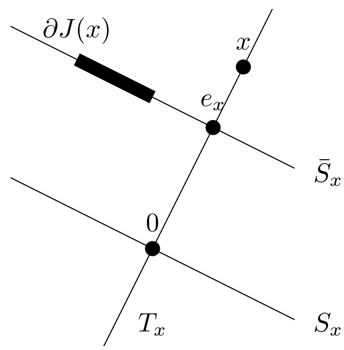

FIG. 1: Illustration of the geometrical elements (S_x, T_x, e_x), in the particular case where $x \in T_x$, for instance when J is a gauge.
In general $e_x \notin \partial J(x)$, which is the situation displayed on Figure 1.

To illustrate these definitions, we now give the examples of the ℓ^1-ℓ^2 and the ℓ^∞ norms. A more comprehensive treatment is provided in Section 7 which is completely devoted to examples.

EXAMPLE 1 (ℓ^1-ℓ^2 norm) We consider a uniform disjoint partition $\mathscr{B}$ of $\{1,\ldots,N\}$,

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

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

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

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

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

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

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

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

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

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

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

and

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

Figure 2 shows graphically these definitions for a particular case of $\ell^1 - \ell^2$ norm in $\mathbb{R}^3$.

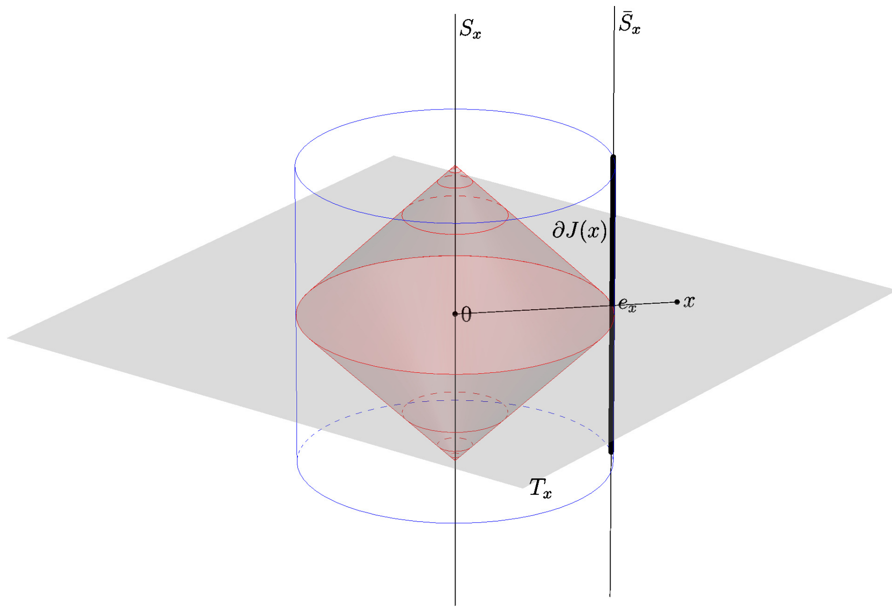

FIG. 2: Illustration of the geometrical elements (S_x, T_x, e_x) for the $\ell^1 - \ell^2$ regularization in dimension 3, for $J(x) = \sqrt{x_1^2 + x_2^2} + |x_3|$ for $x = (x_1, x_2, x_3) \in \mathbb{R}^3$.

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

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

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

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

In turn

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

Figure 3 displays in $\mathbb{R}^3$ these definitions.

3.2 *Decomposability Property*

3.2.1 *The subdifferential 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 a gauge that plays a prominent role in this description.

DEFINITION 4 (Subdifferential Gauge) Let J be a finite-valued convex function. Let $x \in \mathbb{R}^N$ and let $f_x \in \mathrm{ri}\, \partial J(x)$. The *subdifferential gauge* associated to f_x is the gauge $J_{f_x}^{x,\circ} = \gamma_{\partial J(x) - f_x}$.

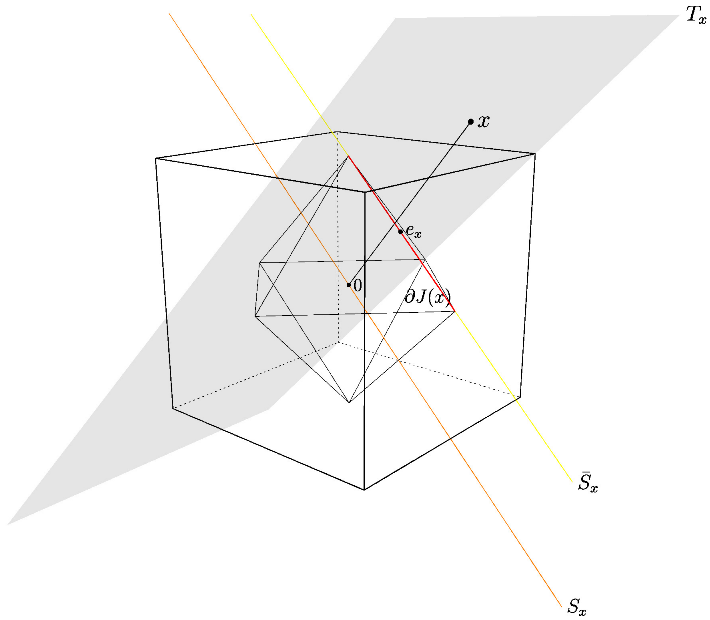

FIG. 3: Illustration of the geometrical elements (S_x, T_x, e_x) for the ℓ^∞ regularization in dimension 3.

Note that for the examples considered so far (ℓ^1, $\ell^1 - \ell^2$ and ℓ^∞ norms), one has $e_x \in \operatorname{ri} \partial J(x)$, so that one can choose $f_x = e_x$ in Definition 4. This is however not the case in general, which makes the introduction of the extra-variable f_x mandatory. In the sequel, it is thus important to remind that $J_{f_x}^{x,\circ}$ actually depends on the particular choice of f_x.

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

PROPOSITION 2 The subdifferential gauge $J_{f_x}^{x,\circ}$ is such that $\operatorname{dom} J_{f_x}^{x,\circ} = S_x$, and is coercive on S_x.

We now turn to the gauge polar to the subdifferential gauge $J_{f_x}^{x} = (J_{f_x}^{x,\circ})^\circ$.The following proposition summarizes its most important properties.

PROPOSITION 3 The gauge $J_{f_x}^{x}$ is such that

(i) It is finite everywhere.

(ii) $J_{f_x}^{x}(d) = J_{f_x}^{x}(d_{S_x}) = \sup_{J_{f_x}^{x,\circ}(\eta_{S_x}) \leqslant 1} \langle \eta_{S_x}, d \rangle$.

(iii) $\operatorname{Ker} J_{f_x}^{x} = T_x$ and $J_{f_x}^{x}$ is coercive on S_x.

3.2.2 *Subdifferential of a gauge.* 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 [HUL01]

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

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 $\left\{ \eta \in \mathbb{R}^N : \ \langle \eta, x \rangle = \gamma_C(x) \right\}$. The special case of $x = 0$ has a much simpler structure; it is the polar set C° from Lemma 3(ii)-(iii), i.e.

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

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 4 Let $J = \gamma_C$ be a finite-valued gauge. Then for $x \in \mathbb{R}^N$, one has

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

PROPOSITION 5 Let $J = \gamma_C$ be a finite-valued gauge. For any $x \in \mathbb{R}^N$, one has

(i) For every $u \in \bar{S}_x$, $J(x) = \langle u, x \rangle$.

(ii) $x \in T_x$.

(iii) The subdifferential gauge $J_{f_x}^{x,\circ}$ reads

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

(iv) The polar of the subdifferential gauge $J_{f_x}^{x}$ reads

$$J_{f_x}^{x}(d) = J(d_{S_x}) - \langle f_{S_x}, d_{S_x} \rangle .$$

We draw the attention of the reader to the fact that $J^\circ, J_{f_x}^{x,\circ}$ and $J_{f_x}^{x}$ are not the same function. The first one is the polar of J, the second one is the subdifferential gauge and the third one is the polar of the subdifferential gauge.

3.2.3 *Decomposability of the subdifferential.* Piecing together the above ingredients yields a fundamental pointwise decomposition of the subdifferential of the regularizer J.

THEOREM 1 (Decomposability) Let J be a convex function. Let $x \in \mathbb{R}^N$ and $f_x \in \operatorname{ri} \partial J(x)$. Then the subdifferential of J at x reads

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

The chosen terminology of "decomposability" appears quite natural in view of the splitting of the subdifferential entailed by the two orthogonal subspaces T_x and S_x. The terminology ($\mathscr{U} - \mathscr{V}$) decomposition is also used in the seminal work of Lemaréchal et al. [LOS00]. The same wording is also employed by Candés and Recht in their paper [CR12], in which the subdifferential exhibits a similar property, but specialized to norms. The decomposability condition used by Negahban et al. [NRWY10], is related to that of [CR12], but is different (see our discussion in the introduction). In fact, it turns out that decomposability is a fundamental properties of the subdifferential of any convex function, and that it should not be a prior hypothesis for our analysis.

This decomposability property is at the heart of our results, because it enables to check whether some vector η satisfies $\eta \in \operatorname{ri}(\partial J(x))$ (see Theorem 3) and also to quantify how far is η from the relative boundary of $\partial J(x)$ (see Theorem 6).

Let us derive the subdifferential gauge for a smooth function and for the the illustrative example of the ℓ^∞ norm. The case of the $\ell^1 - \ell^2$ norm is detailed in Section 3.3.

EXAMPLE 3 (Differentiable convex function) Let J be a convex function which is everywhere differentiable. Then $\partial J(x) = \{\nabla J(x)\}$. It is clear that $S_x = \{0\}$, and thus $T_x = \mathbb{R}^N$ and $e_x = f_x = \nabla J(x)$. Moreover, $J_{f_x}^{x,\circ}(\eta) = \gamma_{\{0\}}(\eta) = \sigma_{\mathbb{R}^N}(\eta) = \iota_0(\eta)$.

EXAMPLE 4 (ℓ^∞ norm) Recall from Section 3.1 that for $J = \|\cdot\|_\infty$, $f_x = e_x = s/|I|$, with $s_{(I)} = \mathrm{sign}(x_{(I)})$, and $s_{(I^c)} = 0$. Let $\mathscr{K}_x = \partial J(x) - e_x$. It can be straightforwardly shown that in this case,

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

This is rewritten as

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

Thus the subdifferential gauge reads

$$J^{x,\circ}_{f_x}(\eta) = \gamma_{\mathscr{K}_x}(\eta) = \max(\gamma_{S_x}(\eta), \gamma_{\mathscr{K}'_x}(\eta)).$$

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

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

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

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

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

PROPOSITION 6 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 $(\mathscr{P}_\lambda(y))$ if, and only if,

$$\Phi_T^*(y - \Phi x) = \lambda e_x \quad \text{and} \quad J^{x,\circ}_{f_x}(\lambda^{-1} \Phi_S^*(y - \Phi x) - \mathrm{P}_S(f_x)) \leqslant 1.$$

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

$$\Phi_T^* \alpha = e_x \quad \text{and} \quad J^{x,\circ}_{f_x}(\Phi_S^* \alpha - \mathrm{P}_S(f_x)) \leqslant 1.$$

3.3 *Strong Gauge*

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

DEFINITION 5 A finite-valued 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 7 Let J be a finite-valued gauge. Then, J is separable w.r.t. to $T = S^\perp$ if, and only if its polar J° satisfies

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

The decomposability of $\partial J(x)$ as described in Theorem 1 depends on the particular choice of the map $x \mapsto f_x \in \mathrm{ri}\,\partial J(x)$. An interesting situation is encountered when $e_x \in \mathrm{ri}\,\partial J(x)$, 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 subspace T, we denote $\widetilde{T}$ the set of vectors sharing the same T,

$$\widetilde{T} = \left\{x \in \mathbb{R}^N : \; T_x = T\right\}.$$

Using positive homogeneity, it is easy to show that $T_{\rho x} = T_x$ and $e_{\rho x} = e_x$ $\forall \rho > 0$, see Proposition 5(i). Thus $\widetilde{T}$ is a non-empty cone which is contained in T by Proposition 5(ii).

DEFINITION 6 (Strong Gauge) A *strong gauge* on T is a finite-valued gauge J such that

1. For every $x \in \widetilde{T}$, $e_x \in \mathrm{ri}\,\partial J(x)$.
2. J is separable with respect to T.

Moreover, if J is a norm, we say that J is a *strong norm* if it is a norm and a strong gauge.

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

PROPOSITION 7 Let J be a strong gauge on T_x. Then, for any $x \in \widetilde{T}$, the subdifferential of J at x reads

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

When J is in addition a norm, this coincides exactly with the decomposability definition of [CR12]. Note however that the last part of assertion (ii) in Proposition 3 is an intrinsic property of the polar of the subdifferential gauge, while it is stated as an assumption in [CR12].

EXAMPLE 5 (ℓ^1-ℓ^2 norm) Recall the notations of this example in Section 3.1. Since $e_x = (\mathcal{N}(x_b))_{b\in\mathscr{B}} \in \mathrm{ri}\,\partial J(x)$, and the ℓ^1-ℓ^2 norm is separable, it is a strong norm according to Definition 6. Thus, its subdifferential at x reads

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

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

4. Uniqueness

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

In the case of ℓ^1-norm, [DH01] has proved the following result.

FACT 2 Let x be a solution of $(\mathscr{P}_\lambda(y))$ (resp. a feasible point of $(\mathscr{P}_0(y))$). Denote $I = \mathrm{supp}(x)$ and $s = \mathrm{sign}(x)$. If the *Strong Null Space Property* holds

$$\forall \delta \in \mathrm{Ker}(\Phi) \setminus \{0\}, \quad \langle s_{(I)}, \delta_{(I)} \rangle < \|\delta_{(I^c)}\|_1, \tag{NSP$^{\mathrm{S}}$}$$

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

In the following, we derive a similar statement for any convex function, which will allow us to obtain uniqueness condition.

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

LEMMA 8 Let x, x' be two solutions of ($\mathscr{P}_\lambda(y)$). Then,

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

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

4.1 *The Strong Null Space Property*

The following theorem gives a sufficient condition to ensure uniqueness of the solution to ($\mathscr{P}_\lambda(y)$) and ($\mathscr{P}_0(y)$), that we coin *Strong Null Space Property*. This condition is a generalization of the Null Space Property introduced in [DH01] and popular in ℓ^1 regularization.

THEOREM 2 Let J be a finite-valued convex function. Let x be a solution of ($\mathscr{P}_\lambda(y)$) (resp. a feasible point of ($\mathscr{P}_0(y)$)) and let $f_x \in \mathrm{ri}(\partial J(x))$. Denote $T = S^\perp = T_x$ the associated model subspace. If the *Strong Null Space Property* holds

$$\forall \delta \in \mathrm{Ker}(\Phi) \setminus \{0\}, \quad \langle e_x, \delta_T \rangle + \langle \mathrm{P}_S(f_x), \delta_S \rangle < J^x_{f_x}(-\delta_S), \tag{NSP$^\mathrm{S}$}$$

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

This result reduces to the one proved in [FPV+13] when J is a strong norm, i.e. decomposable in the sense of [CR12], pre-composed by a linear operator. Note that when specializing (NSP$^\mathrm{S}$) to a strong gauge J, it reads

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

4.2 *Dual Certificates*

In this section we derive from (NSP$^\mathrm{S}$) 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

$$\mathrm{Ker}(\Phi) \cap T = \{0\}. \tag{$\mathscr{C}_T$}$$

We can derive from Theorem 2 the following corollary.

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

Piecing together Proposition 6 and Corollary 1, one can build a particular dual certificate for ($\mathscr{P}_\lambda(y)$), and then state a sufficient uniqueness explicitly in terms of the decomposable structure of the subdifferential of the regularizer J.

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

(i) If

$$\Phi_T^*(y - \Phi x) = \lambda e_x, \tag{4.1}$$
$$J_{f_x}^{x,\circ}\left(\lambda^{-1}\Phi_S^*(y - \Phi x) - \mathrm{P}_S(f_x)\right) < 1. \tag{4.2}$$

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

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

$$\Phi_T^*\alpha = e_x \quad \text{and} \quad J_{f_x}^{x,\circ}(\Phi_S^*\alpha - \mathrm{P}_S(f_x)) < 1,$$

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

5. Partly Smooth Functions Relative to a Subspace

Until now, except of being convex and finite-valued (i.e. full domain), no other assumption was imposed on the regularizer J. But, toward the goal of studying robust recovery by solving $(\mathscr{P}_\lambda(y))$, more will be needed. This is the main reason underlying the introduction of a subclass of finite-valued convex functions J for which the mappings $x \mapsto e_x$, $x \mapsto \mathrm{P}_{S_x}(f_x)$ and $x \mapsto J_{f_x}^\circ$ exhibit local regularity, in some sense to be precized shortly (see Definition 8).

5.1 *Partly Smooth Functions*

The notion of “partly smooth” functions [Lew02] unifies many non-smooth functions known in the literature. Partial smoothness (as well as identifiable surfaces [Wri93]) captures essential features of the geometry of non-smoothness which are along the so-called “active/identifiable manifold”. Loosely speaking, a partly smooth function behaves smoothly as we move on the partial smoothness manifold, and sharply if we move normal to the manifold. In fact, the behaviour of the function and of its minimizers (or critical points) depend essentially on its restriction to this manifold, hence offering a powerful framework for sensitivity analysis theory. In particular, critical points of partly smooth functions move stably on the manifold as the function undergoes small perturbations [Lew02, LZ13].

Specialized to finite-valued convex functions, the definition of partly smooth functions reads as follows.

DEFINITION 7 A finite-valued convex function J is said to be *partly smooth* at x relative to a set $\mathscr{M} \subseteq \mathbb{R}^N$ if

1. **Smoothness.** $\mathscr{M}$ is a C^2-manifold around x and J restricted to $\mathscr{M}$ is C^2 around x.
2. **Sharpness.** The tangent space of $\mathscr{M}$ at x is the model space T_x,
$$\mathscr{T}_{\mathscr{M}}(x) = T_x.$$
3. **Continuity.** The set-valued mapping ∂J is continuous at x relative to $\mathscr{M}$.

The manifold $\mathscr{M}$ is coined a *model manifold* of $x \in \mathbb{R}^N$. J is said to be *partly smooth relative to a set $\mathscr{M}$* if $\mathscr{M}$ is a manifold and J is partly smooth at each point $x \in \mathscr{M}$ relative to $\mathscr{M}$. If J is partly smooth and J is a strong gauge, we say that J is *strongly partly smooth*.

Since J is proper convex and finite-valued, the subdifferential $\partial J(x)$ is everywhere non-empty, compact and convex. Therefore, by [RW98, Corollary 8.11 and Proposition 8.12], the Clarke regularity property [Lew02, Definition 2.7(ii)] is automatically verified. In view of [Lew02, Proposition 2.4(i)-(iii)], our sharpness property is equivalent to that of [Lew02, Definition 2.7(iii)]. Obviously, any smooth function $J : \mathbb{R}^N \to \mathbb{R}$ is partly smooth relative to $\mathbb{R}^N$. Moreover, if $\mathscr{M}$ is a manifold around x, the indicator function $\iota_{\mathscr{M}}$ is partly smooth at x relative to $\mathscr{M}$. Remark that in the previous definition, $\mathscr{M}$ needs only to be defined locally around x, and it can be shown to be locally unique, see [HL04, Corollary 4.2]. Hence the notation $\mathscr{M}$ is unambiguous and we can say that $\mathscr{M}$ is *the* model manifold.

5.2 *Partial Smoothness Relative to a Subspace*

Many of the partly smooth functions considered in the literature are associated to linear subspaces, i.e. in which case the model subspace is the model manifold $\mathscr{M} = T_x$ (see the sharpness property). This class of functions, coined partly smooth functions relative to a subspace, encompasses most of the popular regularizers in signal/image processing, machine learning and statistics. As we will see, ℓ^1, $\ell^1 - \ell^2$, ℓ^∞ norms, their composition by a linear operator, and/or positive combinations of them, to name a few, are partly smooth relative to a subspace. However, this family of regularizers does not include the nuclear norm, whose model manifold is obviously not linear (set of fixed rank matrices). In the reminder of the paper, we focus our attention on the class of regularizers J which are finite-valued convex and partly smooth at x relative to T_x.

In order to derive quantitative stability bounds in Section 6, it is important to quantify precisely the local regularity of the mappings $x \mapsto e_x$, $x \mapsto \mathrm{P}_{S_x}(f_x)$ and $x \mapsto J^{x,\circ}_{f_x}$. This is formalized in the following definition.

DEFINITION 8 Let Γ be any gauge which is finite and coercive on T_x for $x \in \mathbb{R}^N$. Let f be any mapping

$$f : \begin{cases} T_x & \to \mathbb{R}^N \\ \tilde{x} & \mapsto f_{\tilde{x}} \in \mathrm{ri}\, \partial J(\tilde{x}). \end{cases} \tag{5.1}$$

For $(\nu_x, \mu_x, \tau_x, \xi_x) \in \mathbb{R}^4_+$, we denote

$$J \in \mathrm{PSFL}_x(\Gamma, f_x, \nu_x, \mu_x, \tau_x, \xi_x)$$

if J is a finite-valued convex and partly smooth functions at x relative to T_x such that

$$\forall x' \in T_x \quad \text{and} \quad \Gamma(x - x') \leqslant \nu_x \Longrightarrow T_x = T_{x'} \tag{5.2}$$

and for every $x' \in T_x$ with $\Gamma(x - x') < \nu_x$, one has

$$\Gamma(e_x - e_{x'}) \leqslant \mu_x \Gamma(x - x'), \tag{5.3}$$

$$J^{x,\circ}_{f_x}(\mathrm{P}_S(f_x - f_{x'})) \leqslant \tau_x \Gamma(x - x'), \tag{5.4}$$

$$\sup_{\substack{u \in S \\ u \neq 0}} \frac{J^{x',\circ}_{f_{x'}}(u) - J^{x,\circ}_{f_x}(u)}{J^{x,\circ}_{f_x}(u)} \leqslant \xi_x \Gamma(x - x'). \tag{5.5}$$

The following theorem shows that these regularity conditions should really be interpreted as quantitative Lipschitz bounds on the variation of the subdifferential ∂J.

THEOREM 4 Let J be a partly smooth function at x relative to T_x, and assume that $\partial J : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ is Lipschitz-continuous around x relative to T_x. Then for any gauge Γ which is finite and coercive on T_x, and for any Lipschitz map f of the form (5.1), there exists $(\nu_x, \mu_x, \tau_x, \xi_x) \in \mathbb{R}_+^4$ such that $J \in \mathrm{PSFL}_x(\Gamma, f_x, \nu_x, \mu_x, \tau_x, \xi_x)$. Moreover, there always exists such a Lipschitz mapping f.

5.3 *Operations Preserving Partial Smoothness Relative to a Subspace*

The set PSFL_x is closed under addition and pre-composition by a linear operator.

5.3.1 *Addition.* The following proposition determines the model subspace and the subdifferential gauge of the sum of two functions

$$H = J + G$$

in terms of those associated to J and G.

PROPOSITION 8 Let J and G be two finite-valued convex functions. 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) $T^H = T^J \cap T^G$, or equivalently $S^H = (T^H)^\perp = \mathrm{span}\left(S^J \cup S^G\right)$.

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

(iii) Moreover, let $J^{x,\circ}_{f_x^J}$ and $G^{x,\circ}_{f_x^G}$ denote the subdifferential gauges for the pairs $(J, f_x^J \in \mathrm{ri}\, \partial J(x))$ and $(G, f_x^G \in \mathrm{ri}\, \partial G(x))$, correspondingly. Then, for the particular choice of

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

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

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

Armed with this result, we show the following.

PROPOSITION 9 Let $x \in \mathbb{R}^N$. Suppose that

$$J \in \mathrm{PSFL}_x(\Gamma^J, f_x^J, \nu_x^J, \mu_x^J, \tau_x^J, \xi_x^J) \quad \text{and} \quad G \in \mathrm{PSFL}_x(\Gamma^G, f_x^G, \nu_x^G, \mu_x^G, \tau_x^G, \xi_x^G).$$

Then, for the choice $f_x^H = f_x^J + f_x^G$ and $\Gamma^H = \max(\Gamma^J, \Gamma^G)$, we have

$$H = J + G \in \mathrm{PSFL}_x(\Gamma^H, f_x^H, \nu_x^H, \mu_x^H, \tau_x^H, \xi_x^H)$$

with

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

5.3.2 *Smooth perturbation.* It is common in the literature to find regularizers of the form $J_\varepsilon(x) = J(x) + \frac{\varepsilon}{2}\|x\|_2^2$, such as the Elastic net [ZH05] . More generally, we consider any smooth perturbation of J. The following is a straightforward consequence of Proposition 8.

COROLLARY 2 Let J be a finite-valued convex function, $x \in \mathbb{R}^N$ and G a convex function which is differentiable at x. Then,

$$T_x^{J+G} = T^J \quad \text{and} \quad e_x^{J+G} = e_x^J + P_{T_x^J} \nabla G(x).$$

Moreover, for the particular choice of

$$f_x^{J+G} = f_x^J + \nabla G(x),$$

we have $f_x^{J+G} \in \mathrm{ri}(J+G)(x)$ and for a given $\eta \in S_x^J$, the subdifferential gauge of $J+G$ reads

$$(J+G)^{x,\circ}_{f_x^{J+G},x}(\eta) = J^{x,\circ}_{f_x^J,x}(\eta).$$

Hence, the model subspace T_x and the subdifferential gauge are insensitive to smooth perturbations. Combining Proposition 9 and Corollary 2 yields the partial smoothness Lipschitz constants of smooth perturbation.

COROLLARY 3 Let $x \in \mathbb{R}^N$. Suppose that $J \in \mathrm{PSFL}_x(\Gamma^J, f_x^J, \nu_x^J, \mu_x^J, \tau_x^J, \xi_x^J)$, that G is C^2 on $\mathbb{R}^N$ with a β-Lipschitz gradient. Then for the choice $f_x^H = f_x^J + \nabla G(x)$ and $\Gamma^H = \max(\Gamma^J, \|\cdot\|)$, $H = J + G \in \mathrm{PSFL}_x(\Gamma^H, f_x^H, \nu_x^H, \mu_x^H, \tau_x^H, \xi_x^H)$ with

$$\nu_x^H = \nu_x^J, \quad \mu_x^H = \mu_x^J\, |||\mathrm{P}_{T^J}|||_{\Gamma^J \to \Gamma^H} + \beta\, |||\mathrm{P}_{T^J}|||_{\ell^2 \to \Gamma^H},$$
$$\tau_x^H = \tau_x^J, \quad \xi_x^H = \xi_x^J.$$

5.3.3 *Pre-composition by a Linear Operator.* Convex functions of the form $J_0 \circ D^*$, where J_0 is a finite-valued convex function, 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 J_0 at D^*x. The following proposition details the decomposability structure of analysis-type regularizers.

PROPOSITION 10 Let J_0 be a convex finite-valued function. Then the subdifferential of $J = J_0 \circ D^*$ has the decomposability property with

(i) $T = \mathrm{Ker}(D_{S_0}^*)$, or equivalently $S = \mathrm{Im}(D_{S_0})$.

(ii) $e = D_T e_0$.

(iii) Moreover, let $J^{D^*x,\circ}_{0,f_{0,D^*x}}$ denote the subdifferential gauge for the pair $(J_0, f_{0,D^*x} \in \mathrm{ri}\,\partial J_0(x))$. Then, for the particular choice of

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

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

$$J^{x,\circ}_{f_x}(\eta) = \inf_{z \in \mathrm{Ker}(D_{S_0})} J^{D^*x,\circ}_{0,f_{0,D^*x}}(D_{S_0}^+ \eta + z) .$$

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

Capitalizing on these properties, we now establish the following.

PROPOSITION 11 Let $x \in \mathbb{R}^N$ and $u = D^* x$. Suppose that $J_0 \in \mathrm{PSFL}_u(\Gamma_0, f_{0,u}, \nu_{0,u}, \mu_{0,u}, \tau_{0,u}, \xi_{0,u})$. Then with the choice $f_x = D f_{0,u}$ and Γ any finite-valued coercive gauge on T, $J = J_0 \circ D^* \in \mathrm{PSFL}_x(\Gamma, f_x, \nu_x, \mu_x, \tau_x, \xi_x)$, with

$$
\begin{aligned}
\nu_x &= \frac{1}{|||D^*|||_{\Gamma \to \Gamma_0}} \nu_{0,u} \\
\mu_x &= \mu_{0,u} |||D_T|||_{\Gamma \to \Gamma_0} |||D^*|||_{\Gamma \to \Gamma_0} \\
\tau_x &= \left(\tau_{0,u} \left|\left|\left| D_{S_0}^{+} D_S \right|\right|\right|_{J^{u,\circ}_{0,f_{0,u}} \to J^{u,\circ}_{0,f_{0,u}}} + \mu_{0,u} \left|\left|\left| D_{S_0}^{+} D_S \right|\right|\right|_{\Gamma_0 \to J^{u,\circ}_{0,f_{0,u}}} \right) |||D^*|||_{\Gamma \to \Gamma_0} \\
\xi_x &= \xi_{0,u} |||D^*|||_{\Gamma \to \Gamma_0} \ .
\end{aligned}
$$

6. Exact Model Selection and Identifiability

In this section, we state our main recovery guarantee. This result asserts that under appropriate conditions, and for small enough noise, $(\mathscr{P}_\lambda(y))$ with a partly smooth function J at x_0 relative to the subspace T_{x_0} has a unique solution $x^\star$, and moreover, its model subspace equals that of x_0, i.e. $T_{x^\star} = T_{x_0}$. Put differently, provided that the noise is sufficiently small, regularization by J is able to stably recover the correct model subspace underlying x_0.

6.1 *Linearized Precertificate*

Let us first introduce the definition of the linearized precertificate.

DEFINITION 9 The *linearized precertificate* α_F for $x \in \mathbb{R}^N$ is defined by

$$
\alpha_F = \underset{\Phi_{T_x}^* \alpha = e_x}{\operatorname{argmin}} \ \|\alpha\|.
$$

The subscript F is used as a salute to J.-J. Fuchs [Fuc04] who first considered this vector as a dual certificate for ℓ^1 minimization. The intuition behind it is well-understood if one realizes that the existence of a dual certificate α is equivalent to $\eta = \Phi^* \alpha$ for some α such that $\eta_T = e_x$ and $J_{f_x}^{x,\circ}(\eta_S - P_S f_x) \leqslant 1$. Dropping the last constraint, and choosing the minimal ℓ^2-norm solution to the first constraint recovers the definition of α_F.

A convenient property of this vector, is that under the restricted injectivity condition, it has a closed form expression.

LEMMA 9 Let $x \in \mathbb{R}^N$ and suppose that $(\mathscr{C}_T)$ is verified with $T = T_x$. Then α_F is well-defined and

$$
\alpha_F = \Phi_{T_x}^{+,*} e_x.
$$

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

DEFINITION 10 For $x \in \mathbb{R}^N$ such that $(\mathscr{C}_{T_x})$ with $T = T_x$ holds, we define the *Irrepresentability Criterion* at x as

$$
\mathbf{IC}(x) = J_{f_x}^{x,\circ}(\Phi_{S_x}^* \Phi_{T_x}^{+,*} e_x - \mathrm{P}_{S_x} f_x).
$$

A fundamental remark is that $\mathbf{IC}(x) < 1$ is the analytical equivalent to the topological non-degeneracy condition $\Phi^* \alpha_F \in \operatorname{ri} \partial J(x)$. Note that if J is a strong gauge on T, then it reads $\mathbf{IC}(x) = J^\circ(\Phi_{S_x}^* \Phi_{T_x}^{+,*} e_x)$, see Proposition 7. The Irrepresentability Criterion clearly brings into play the promoted subspace T_x and the interaction between the restriction of Φ to T_x and S_x. It is a generalization of the irrepresentability condition that has been studied in the literature for some popular regularizers, including the ℓ^1-norm [Fuc04], analysis-ℓ^1 [VPDF13], and ℓ^1-ℓ^2 [Bac08a]. See Section 7 for a comprehensive discussion.

6.2 *Exact Model Selection*

We begin with the noiseless case, i.e. $w = 0$ in (1.1). In fact, in this setting, $\mathbf{IC}(x_0) < 1$ is a sufficient condition for identifiability without any any other particular assumption on the finite-valued convex function J, such as partial smoothness. By identifiability, we mean the fact that x_0 is the unique solution of $(\mathscr{P}_0(y))$.

THEOREM 5 Let $x_0 \in \mathbb{R}^N$ and $T = T_{x_0}$. We assume that $(\mathscr{C}_T)$ holds and $\mathbf{IC}(x_0) < 1$. Then x_0 is the unique solution of $(\mathscr{P}_0(y))$.

It turns out that even in presence of noise in the measurements y according to (1.1), condition $\mathbf{IC}(x_0) < 1$ is also sufficient for $(\mathscr{P}_\lambda(y))$ with $PSFL_{x_0}$ regularizer to stably recover the model subspace underlying x_0. This is stated in the following theorem.

THEOREM 6 Let $x_0 \in \mathbb{R}^N$ and $T = T_{x_0}$. Suppose that $J \in \mathrm{PSFL}_{x_0}(\Gamma, \nu_{x_0}, \mu_{x_0}, \tau_{x_0}, \xi_{x_0})$. Assume that $(\mathscr{C}_T)$ 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

$$\frac{A_T}{1 - \mathbf{IC}(x_0)} \|w\| \leqslant \lambda \leqslant \nu_{x_0} \min\big(B_T, C(x_0)\big) \tag{6.1}$$

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

$$\|x_0 - x^\star\| = O\Big(\max(\|w\|, \lambda)\Big).$$

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 (6.1). 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 (6.1) to be non-empty, the noise-to-signal level $\|w\|/\nu_{x_0}$ should be small enough, i.e.

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

See the illustrative examples detailed in Section 7 for concrete expressions of the parameter ν_{x_0} and how it relates to a minimal signal level.

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

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

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

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 local geometry of the partly smooth regularizer 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.

It is worth noting that this analysis is in some sense sharp following the argument in [VPF14, Proposition 1]. The only case not covered by our analysis is when $\mathbf{IC}(x) = 1$.

7. Examples of Partly Smooth Functions Relative to a Subspace

7.1 *Synthesis ℓ^1 Sparsity*

The regularized problem ($\mathscr{P}_\lambda(y)$) with $J(x) = \|x\|_1 = \sum_{i=1}^N |x_i|$ promotes sparse solutions. It goes by the name of Lasso [Tib96] in the statistical literature, and Basis Pursuit DeNoising (or Basis Pursuit in the noiseless case) [CDS99] in signal processing.

7.1.1 *Structure of the ℓ^1 norm.* The norm $J(x) = \|x\|_1$ is a symmetric (finite-valued) strong gauge. More precisely, we have the following result.

PROPOSITION 12 $J = \|\cdot\|_1$ is a symmetric strong gauge with

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

where $I = I(x) = \{i : x_i \neq 0\}$. Moreover, it is partly smooth relative to a subspace with

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

7.1.2 *Relation to previous works.* 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 irrepresentability criterion was introduced in [Fuc04]. Let $s \in \{-1, 0, +1\}^N$ and I its support. Suppose that $\Phi_{(I)}$ has full column rank, which is precisely ($\mathscr{C}_T$) in this case. The synthesis irrepresentability criterion $\mathbf{IC}_{\ell^1}$ 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 12, one immediately recognizes that $\mathbf{IC}_{\ell^1}(\operatorname{sign}(x)) = \mathbf{IC}(x)$. The condition $\mathbf{IC}_{\ell^1}(\operatorname{sign}(x)) < 1$, also known as the irrepresentability condition in the statistical literature, was proposed [Fuc04] 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 6 in this case ensures recovery of the support pattern.

7.2 *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 $D = \mathrm{Id}$. Popular examples of analysis-type ℓ^1 semi-norms include for instance the discrete (anisotropic) total variation [ROF92], the Fused Lasso [TSR$^+$04] and shift invariant wavelets [SWB$^+$04].

7.2.1 *Structure of the analysis ℓ^1 semi-norm.* The semi-norm $J(x) = \|D^* x\|_1$ is a symmetric partly smooth function relative to a subspace. This is formalized in the following proposition whose proof is a straightforward application of Proposition 10, Proposition 11 and Proposition 12.

PROPOSITION 13 $J = \|D^* \cdot\|_1$ is a symmetric (finite-valued) gauge with

$$T_x = \mathrm{Ker}(D^*_{(I^c))}) = \left\{\eta \in \mathbb{R}^N : \forall j \notin I, \langle d_j, \eta_j \rangle = 0\right\}, \quad S_x = \mathrm{Im}(D_{I^c}),$$
$$e_x = \mathrm{P}_{\mathrm{Ker}(D^*_{I^c})} D \operatorname{sign}(D^* x), \quad f_x = D \operatorname{sign}(D^* x),$$
$$J^{x,\circ}_{f_x}(\eta) = \inf_{z \in \mathrm{Ker}(D_{(I^c)})} \|D^+_{(I^c)} \eta + z\|_\infty, \quad \text{for} \quad \eta \in S_x \ ,$$

where $I = I(x) = \{i : \langle d_i, x_i \rangle \neq 0\}$. Moreover, it is partly smooth relative to a subspace with parameters

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

7.2.2 *Relation to previous works.* Some insights on the relation and distinction between synthesis- and analysis-based sparsity regularizations were first given in [EMR07]. 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 [VPDF13] for a comprehensive review. Among such a work, the authors in [NDEG13] propose a null space property for identifiability in the noiseless case and in [KRZ14] one can find results in the gaussian setting. The most relevant work to ours here is that of [VPDF13], where the authors prove exact robust recovery of the support and sign patterns under conditions that are a specialization of those in Theorem 6.

More precisely, let I be the support of $D^* x_0$, and s its sign vector. Denote $T = T_{x_0} = S^\perp = \mathrm{Ker}(D^*_{I^c})$, $e_{x_0} = \operatorname{sign}(D^* x_0) = s$, $e = e_{x_0} = \mathrm{P}_T Ds$, $f = f_{x_0} = Ds$. From Definition 10 and Proposition 13, the criterion $\mathbf{IC}(x_0)$ in this case takes the form

$$\begin{aligned}
\mathbf{IC}(x_0) &= J^{x,\circ}_{f_x}(\Phi_S^* \Phi_T^{+,*} \mathrm{P}_T Ds - \mathrm{P}_S Ds) \\
&= \inf_{z \in \mathrm{Ker}(D_{(I^c)})} \|D^+_{(I^c)} \left(\Phi_S^* \Phi_T^{+,*} \mathrm{P}_T - \mathrm{P}_S\right) Ds + z\|_\infty \\
&= \inf_{z \in \mathrm{Ker}(D_{(I^c)})} \|D^+_{(I^c)} \left((\mathrm{Id} - \mathrm{P}_T)\Phi^* \Phi \, \mathrm{P}_T (\Phi_T^* \Phi_T)^{-1} \mathrm{P}_T - \mathrm{P}_S\right) Ds + z\|_\infty \\
&= \inf_{z \in \mathrm{Ker}(D_{(I^c)})} \|D^+_{(I^c)} \left(\Phi^* \Phi \, \mathrm{P}_T (\Phi_T^* \Phi_T)^{-1} \mathrm{P}_T - (\mathrm{P}_T + \mathrm{P}_S)\right) Ds + z\|_\infty \\
&= \inf_{z \in \mathrm{Ker}(D_{(I^c)})} \|D^+_{(I^c)} \left(\Phi^* \Phi \, \mathrm{P}_T (\Phi_T^* \Phi_T)^{-1} \mathrm{P}_T - \mathrm{Id}\right) D_{(I)} s_{(I)} + z\|_\infty \ .
\end{aligned}$$

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

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

where $A^{[I^c]} = U(U^* \Phi^* \Phi U)^{-1} U^*$. We recover exactly the expression of the $\mathbf{IC}_{\ell^1-D}$ introduced in [VPDF13].

7.3 ℓ^∞ *Antisparsity Regularization*

Regularization by the ℓ^∞-norm corresponds to taking $J(x) = \|x\|_\infty = \max_{1 \leqslant i \leqslant N} |x_i|$. This regularizer promotes flat solutions. It plays a prominent role in a variety of applications including approximate nearest neighbor search [JFF12] or vector quantization [LV10]; see also [SYB12] and references therein.

7.3.1 *Structure of the ℓ^∞-norm.* The norm $J(x) = \|x\|_\infty$ is a symmetric partly smooth function relative to a subspace, 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 14 $J = \|\cdot\|_\infty$ is a symmetric (finite-valued) gauge with

$$S_x = \left\{\eta : \ \eta_{(I^c)} = 0 \quad \text{and} \quad \langle \eta_{(I)}, s_{(I)} \rangle = 0\right\}, \quad T_x = \left\{\alpha : \ \alpha_{(I)} = \rho s_{(I)} \quad \text{for} \quad \rho \in \mathbb{R}\right\},$$
$$e_x = \frac{s}{|I|}, \quad f_x = e_x, \quad J^{x,\circ}_{f_x}(\eta) = \max_{i \in I} \, (-|I| s_i \eta_i)_+ \quad \text{for} \quad \eta \in S_x \ ,$$

where $s = \mathrm{sign}(x)$ and $I = I(x) = \{i : \ |x_i| = \|x\|_\infty\}$. Moreover, it is partly smooth relative to a subspace with

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

7.3.2 *Relation to previous work.* In the noiseless case, i.e. $(\mathscr{P}_0(y))$ with $J = \|\cdot\|_\infty$, theoretical analysis of ℓ^∞-regularization goes back to the 70's through the work of [Cad71]. [LV10] 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 $(\mathscr{P}_0(y))$ is derived in [MR11]. The work of [DT10] analyzes recovery guarantees by ℓ^∞-regularization in a noiseless random sensing setting.

The authors in [SYB12] analyzed the properties of solutions obtained from a constrained form of $(\mathscr{P}_\lambda(y))$ with $J = \|\cdot\|_\infty$. In particular, they improved and generalized the bound of [LV10] on the ℓ^∞ of the solution.

The work of [Bac10, OB12] 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 7.5.

7.4 *Group Sparsity Regularization*

Let's recall from Section 3.1 that $\mathscr{B}$ is a uniform disjoint partition of $\{1, \cdots, N\}$,

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

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

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

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 [Bak99, YL05]. The corresponding regularized problem $(\mathscr{P}_\lambda(y))$ is known as the group Lasso.

7.4.1 *Structure of the ℓ^1-ℓ^2 norm.* The $\ell^1-\ell^2$ norm is a symmetric partly smooth function relative to a subspace.

PROPOSITION 15 The $\ell^1-\ell^2$ norm associated to the partition $\mathscr{B}$ is a symmetric (finite-valued) strong gauge with

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

where $I = I(x) = \{b : x_b \neq 0\}$, and $\mathscr{N}(a) = a/\|a\|$ if $a \neq 0$, and $\mathscr{N}(0) = 0$. Moreover, it is partly smooth relative to a subspace with

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

7.4.2 *Relation to previous work.* Theoretical guarantees of the group Lasso have been investigated by several authors under different performance criteria; see e.g. [YL05, RF08, Bac08a, CH08, LZ09, WH10] to cite only a few. In particular, the author in [Bac08a] studies the asymptotic group selection consistency of the group Lasso in the overdetermined case, under a group irrepresentability condition. This condition also appears in noiseless identifiability in the work of [CR12]. The group irrepresentability 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 15, and assuming that $\Phi_{(I)}$ is full column rank (i.e. $(\mathscr{C}_T)$ is fulfilled), $\mathbf{IC}(x_0)$ reads

$$\mathbf{IC}(x_0) = \left\| \Phi^*_{(I^c)} \Phi^{+,*}_{(I)} \left(\tfrac{x_b}{\|x_b\|}\right)_{b\in I} \right\|_{\infty,2} . \tag{7.1}$$

It is worth mentioning that the discrete isotropic total variation in d-dimension, $d \geqslant 2$, can be viewed as an analysis-type $\ell^1-\ell^2$ semi-norm. Partial smoothness 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 5.3.3.

7.5 *Polyhedral Regularization*

The ℓ^1 and ℓ^∞ norms are special cases of polyhedral priors. 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.

7.5.1 *Structure of polyhedral gauges.* A polyhedral gauge in the H-representation is defined as

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

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 5.3.3 by identifying $D = H$. One can then characterize decomposability and partial smoothness relative to a subspace of J_0 and then invoke Proposition 10 and 11 to derive those of J. This is what we are about to do. In the following, we denote $(a^i)_{1\leqslant i\leqslant N_H}$ the standard basis of $\mathbb{R}^{N_H}$.

PROPOSITION 16 $J_0(u) = \max_{1 \leqslant i \leqslant N_H} (u_i)_+$ is a (finite-valued) gauge and,

- If $u_i \leqslant 0, \forall i \in \{1, \cdots, N_H\}$, then

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

  where

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

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

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

  where

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

Moreover, it is partly smooth relative to a subspace with parameters (assuming $I_+ \neq \emptyset$)

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

7.5.2 *Relation to previous works.* As stated in the case of ℓ^∞-norm, the work of of [Bac10] considers robust recovery with a subclass of polyhedral norms but his notion of support is different from ours. The work [PT12] studies numerically some polyhedral regularizations. Again in a compressed sensing scenario, the work of [CRPW12] 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 [VPF13], 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 finite-valued coercive polyhedral gauges were considered there.

7.6 *A Counter-Example: the Nuclear Norm*

The nuclear norm is the natural extension of ℓ^1 sparsity to matrix-valued data $x \in \mathbb{R}^{N_0 \times N_0}$ (where $N = N_0^2$). We denote $x = V_x \operatorname{diag}(\Lambda_x) U_x^*$ an SVD decomposition of x, where $\Lambda_x \in \mathbb{R}_+^{N_0}$. Note that this can be extended easily to rectangular matrices. The nuclear norm imposes such a sparsity and is defined as

$$J(x) = \|x\|_* = \|\Lambda_x\|_1,$$

see [VPF14] and the reference therein. This norm can be shown to be partly smooth (in the sense of Definition 7) at some x with respect to the set $\mathscr{M} = \{x' : \operatorname{rank}(x) = \operatorname{rank}(x')\}$ that is locally a manifold around x. This manifold is however not a linear space, hence one does not have $\mathscr{M} = T_x$. This shows that the nuclear norm is not in the set PSFL_x of functions that are partly smooth with respect to a subspace (in the sense of Definition 8). In particular, Theorem 6 cannot be applied to this functional.

It is however possible to show that the manifold $\mathscr{M}$ associated to x is stable to small noise perturbation in the observation under the same hypotheses as Theorem 6. This result is proved in [VPF14], which extends the previous result of Bach [Bac08b]. Note however that these proofs do not give explicit stability constants, in contrast to Theorem 6.

8. Case Study: Compressed Sensing with ℓ^∞ Regularization

In this section, based on the generalized irrepresentability condition, we provide a bound for the sampling complexity to guarantee exact and stable recovery of the model subspace T_{x_0} of anti-sparsity minimization from noisy Gaussian measurements.

THEOREM 7 Let x be an arbitrary vector with its saturation support I, its model tangent subspace $T_x = S_x^\perp$ and model vector e_x as defined in Proposition 14. Let $\beta > 1$. For Φ drawn from the standard Gaussian ensemble with

$$Q \geqslant N - |I| + 2\beta |I| \log(|I|/2) ,$$

$\mathbf{IC}(x) < 1$ with probability at least $1 - 2(|I|/2)^{-f(\beta,|I|)}$ where

$$f(\beta,|I|) = \left(\sqrt{\frac{\beta}{2|I|} + \beta - 1} - \sqrt{\frac{\beta}{2|I|}} \right)^2 .$$

The above bound and probability bears some similarities to what we get with ℓ^1 minimization, except that now the probability of success scales in a power of $|I|$ and not N directly. The reason underlying such a similarity is the proof technique usual in compressed sensing-type bounds and the use of the minimal ℓ^2-norm dual certificate.

The map $f(\beta,|I|)$ is an increasing function of $|I|$, so that $\lim_{|I|\to\infty} f(\beta,|I|) = \beta - 1$ and the probability of success increases with increasing size of the saturation support. But this comes at the price of a stronger requirement on the number of measurements.

For the noiseless problem $(\mathscr{P}_0(y))$, it can be shown using arguments based on the statistical dimension [ALMT13] of the descent cone of the ℓ^∞-norm that there is a phase transition exactly at $N - |I|/2$, see also [CRPW12, Proposition 3.12]. The reason is that each face of the descent cone of the hypercube at a point living on its k-dimensional face is the direct sum of a subspace (the subspace parallel to the face), and of an orthant of dimension $N-k$ (up to an isometry). The statistical dimension is then $(N-k)/2 + k = (N+k)/2 = N - |I|/2$, observing that $k = N - |I|$.

9. Conclusion

In this paper, we introduced the notion of partly smooth function relative to a subspace as a generic convex regularization framework, and presented a unified view 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 $(\mathscr{P}_\lambda(y))$ and $(\mathscr{P}_0(y))$, whose by-product is to guarantee exact recovery of the original object x_0 in the noiseless case by solving $(\mathscr{P}_0(y))$. In presence of noise, sufficient sharp 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, $\ell^1-\ell^2$), as well as less popular ones (ℓ^∞, polyhedral). We exemplified the usefulness of this analysis by providing a sampling complexity bound for exact support recovery in ℓ^∞ regularization from Gaussian measurements.

A. Proofs of Section 2

Proof of Lemma 2. (i)-(iii) are obtained from [HUL01, Theorem V.1.2.5]. (iv) is obtained by combining [HUL01, 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 $\operatorname{dom}\gamma_C = \operatorname{aff} C = \operatorname{par} C$ since $0 \in \operatorname{ri} C$. □

Proof of Lemma 3. (i) follows from [Roc96, Theorem 15.1]. (ii) [Roc96, Corollary 15.1.1] or [HUL01, Proposition V.3.2.4]. (iii) [Roc96, Corollary 15.1.2] or [HUL01, Proposition V.3.2.5]. □

Proof of Lemma 4. We have from Lemma 3 and calculus rules on support functions,

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

Thus

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

This yields that

$$\begin{aligned}
\gamma_{C_1+C_2}(x) &= \sigma_{(C_1+C_2)^\circ}(x) \\
&= \sigma_{\sigma_{C_1}(u)+\sigma_{C_2}(u)\leqslant 1}(x) \\
&= \sup_{\sigma_{C_1}(u)+\sigma_{C_2}(u)\leqslant 1} \langle u, x\rangle \\
&= \sup_{\rho\in[0,1]} \sup_{\sigma_{C_1}(u)\leqslant\rho,\sigma_{C_2}(u)\leqslant 1-\rho} \langle u, x\rangle \\
&= \sup_{\rho\in[0,1]} \sigma_{\sigma_{C_1}(u)\leqslant\rho} \overset{+}{\vee} \sigma_{\sigma_{C_2}(u)\leqslant 1-\rho}(x) && \text{[HUL01, Proposition 1.3.2]} \\
&= \sup_{\rho\in[0,1]} \rho\,\sigma_{\sigma_{C_1}(u)\leqslant 1} \overset{+}{\vee} (1-\rho)\sigma_{\sigma_{C_2}(u)\leqslant 1}(x) && \text{Positive homogeneity} \\
&= \sup_{\rho\in[0,1]} \rho\,\sigma_{C_1^\circ} \overset{+}{\vee} (1-\rho)\sigma_{C_2^\circ}(x) && \text{Polarity} \\
&= \sup_{\rho\in[0,1]} \rho\,\gamma_{C_1} \overset{+}{\vee} (1-\rho)\gamma_{C_2}(x) \; , && \text{Lemma 3}
\end{aligned}$$

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 assumptions of the lemma, the objective in the sup inf is a continuous finite concave-convex function[2] on $[0,1]\times\{(x_1,x_2) : \; x_1+x_2 = x\}$. Since the latter sets are non-empty, closed and convex, and

[2] A concave-convex function f on $C\times D$ is a function such that for each $c\in C$, the function $d\mapsto f(c,d)$ is concave, and for each $d\in D$, the function $c\mapsto f(c,d)$ is convex.

$[0,1]$ is obviously bounded, we have from using [Roc96, 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 5. It is immediate to see that $D(C)$ is a compact convex set containing the origin. Moreover, σ_C is finite-valued by compactness of C, and thus $\sigma_C \circ D^*$ is finite-valued. Thus, we have

$$\begin{aligned}\gamma_{(D(C))^\circ} &= \sigma_{D(C)} && \text{Lemma 3} \\ &= \left(\iota_{D(C)}\right)^* && \text{Legendre-Fenchel conjugacy} \\ &= \sigma_C \circ D^* && \text{[HUL01, Theorem X.2.1.1].}\end{aligned} \tag{A.1}$$

Now, recall that by Lemma 3, $\gamma_{C^\circ} = \sigma_C$ which is then finite-valued owing to compactness of C. In view of Lemma 2(iii), this is equivalent to $0 \in \mathrm{int}(C^\circ)$. Therefore we have the qualification condition $\mathrm{Im}(D^*) \cap \mathrm{int}(C^\circ) \neq \emptyset$. We then obtain

$$\begin{aligned}\gamma_{D(C)}(x) &= \sigma_{(D(C))^\circ}(x) && \text{By definition} \\ &= \sigma_{\sigma_C \circ D^*(u) \leqslant 1}(x) && \text{From (A.1)} \\ &= \left(\iota_{\sigma_C(w)\leqslant 1} \circ D^*\right)^*(x) && \text{Legendre-Fenchel conjugacy} \\ &= \inf_v \sigma_{\sigma_C(w)\leqslant 1}(v) \quad \text{s.t.} \quad Dv = x && \text{[HUL01, Theorem X.2.2.3]} \\ &= \inf_{z\in\mathrm{Ker}(D)} \sigma_{\sigma_C(w)\leqslant 1}(D^+x+z) && \text{Change of variable} \\ &= \inf_{z\in\mathrm{Ker}(D)} \gamma_C(D^+x+z) && \text{Lemma 3 .}\end{aligned}$$

□

Proof of Proposition 4. Let $x \in \mathbb{R}^N$. We have

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

where $H = \left\{\eta \in \mathbb{R}^N : \langle \eta, x\rangle = J(x)\right\}$ is the supporting hyperplane of C° at x. By Proposition 5(i), we have

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

which implies that

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

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

Proof of Proposition 5.

(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[3] applied to the gauge J, and using Lemma 3(iii), we have

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

[3]The Fenchel identity states that for a closed function, $f(x) + f^*(s) = \langle s, x\rangle$ if, and only if, $s \in \partial f(x)$.

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) 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 $x \in S_x^\perp = T_x$.

(iii) Since $f_x \in \operatorname{ri} \partial J(x) \subset \bar{S}_x$, Proposition 1 implies that $f_x = \mathrm{P}_{S_x}(f_x) + \mathrm{P}_{T_x}(f_x) = \mathrm{P}_{S_x}(f_x) + e_x$. Hence, using Proposition 4, we get

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

We therefore obtain

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

At this stage, Lemma 4 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 1, $\{-f_x\}^\circ = \{\eta : \langle \eta, f_x \rangle \geqslant -1\}$, which indeed contains the origin as an interior point. Continuing from the last equality, we get using Lemma 3,

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

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

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

Recalling that J° is a finite-valued gauge, hence continuous, the objective in the sup inf fulfills the assumption of the second assertion of Lemma 4, whence we get

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

(iv) Using some calculus rules with support functions and assertion (ii), we have

$$\begin{aligned}
J^x_{f_x}(d) = J^x_{f_x}(d_{S_x}) &= \sigma_{(C^\circ+\{-f_x\})\cap S_x}(d_{S_x}) && \text{By definition of } J^{x,\circ}_{f_x}\\
&= \overline{\mathrm{conv}}\left(\inf(\sigma_{C^\circ+\{-f_x\}}(d_{S_x}), \sigma_{S_x}(d_{S_x}))\right) && \text{[HUL01, Theorem 3.3.3(iii)]}\\
&= \overline{\mathrm{conv}}\left(\inf(\sigma_{C^\circ+\{-f_x\}}(d_{S_x}), \iota_{T_x}(d_{S_x}))\right) && \text{Conjugacy rule on subspaces}\\
&= \sigma_{C^\circ+\{-f_x\}}(d_{S_x}) && d_{S_x}\in S_x = T_x^\perp\\
&= \sigma_{C^\circ}(d_{S_x}) - \langle \mathrm{P}_{S_x}(f_x), d_{S_x}\rangle && \text{[HUL01, Theorem 3.3.3(i)]}\\
&= J(d_{S_x}) - \langle \mathrm{P}_{S_x}(f_x), d_{S_x}\rangle && \text{Lemma 3 and definition of } J \ .
\end{aligned}$$

□

Proof of Lemma 1. To lighten the notation, denote $V = \mathrm{par}\, C$.

(i) [HUL01, Proposition V.2.1.2].

(ii) [HUL01, Proposition V.2.1.3].

(iii) Immediate from the definition and $0 \in C$.

(iv) As $0 \in C$ we have

$$0 \leqslant \sigma_C(d) \leqslant \sigma_{\mathrm{aff}\, C}(d) = \sigma_V(d) \ .$$

Thus $\sigma_C(d) = 0$, $\forall d \in V^\perp$, or equivalently, $V^\perp \subset \mathrm{Ker}\, \sigma_C$, whence we get that $\sigma_C(d) = \sigma_C(d_V)$.

(v) The fact that σ_C is finite-valued is a consequence of (ii) since C is assumed bounded. Now, in view of [HUL01, Theorem V.2.2.3], we have the equivalent characterization

$$0 \in \mathrm{ri}\, C \Leftrightarrow \sigma_C(d) > 0 \quad \forall d \text{ such that } \sigma_C(d) + \sigma_C(-d) > 0.$$

By definition of the support function and closedness of C, $\sigma_C(d) + \sigma_C(-d) > 0$ if and only if there exists two points x and x' in C satisfying $\langle x - x', d\rangle > 0$, or equivalently $d \notin (C-C)^\perp = V^\perp$. We then conclude that $0 \in \mathrm{ri}\, C \Leftrightarrow \sigma_C(d) > 0, \forall d \notin V^\perp$. Combining this with (iv), the claim follows.

□

Proof of Lemma 6. Lipschitz continuity of F on U means that for any pair x, x' in U, we have

$$F(x) \subseteq F(x') + \beta\|x - x'\|\mathbb{B}(0) \quad \text{and} \quad F(x') \subseteq F(x) + \beta\|x - x'\|\mathbb{B}(0) \ ,$$

which in turn is equivalent to

$$\begin{aligned}
\sigma_{F(x')}(u) &\leqslant \sigma_{F(x)+\beta\|x'-x\|\mathbb{B}(0)} = \sigma_{F(x)}(u) + \beta\|x - x'\|\|u\|\\
\sigma_{F(x)}(u) &\leqslant \sigma_{F(x')+\beta\|x'-x\|\mathbb{B}(0)} = \sigma_{F(x')}(u) + \beta\|x - x'\|\|u\| \ ,
\end{aligned}$$

and thus

$$|\sigma_{F(x')}(u) - \sigma_{F(x)}(u)| \leqslant \beta \|x' - x\| \|u\|.$$

By assumption, for any $x \in U$, $F(x)$ is compact, and thus $\sigma_{F(x)}$ is everywhere finite by Lemma 1(ii). Moreover, since $0 \in \mathrm{ri}\, F(x)$, we have from Lemma 1(v) that $\sigma_{F(x)}$ is coercive on $\mathrm{par}(F(x))$. Moreover, $\mathrm{dom}(\gamma_{F(x)}) = \mathrm{par}(F(x))$ and $\gamma_{F(x)}$ is coercive on $\mathrm{par}(F(x))$; see Lemma 2(v). It follows from this coercivity and finiteness that for any $u \in \mathrm{par}(F(x))$, one has

$$\sigma_{F(x)}(u) \leqslant |||\mathrm{Id}|||_{\sigma_{F(x)} \to \gamma_{F(x)}} \gamma_{F(x)}(u) \leqslant \underbrace{\left(\sup_{x \in U} |||\mathrm{Id}|||_{\sigma_{F(x)} \to \gamma_{F(x)}}\right)}_{C_{\sigma \to \gamma}} \gamma_{F(x)}(u) \tag{A.2}$$

$$\gamma_{F(x)}(u) \leqslant |||\mathrm{Id}|||_{\gamma_{F(x)} \to \sigma_{F(x)}} \sigma_{F(x)}(u) \leqslant \underbrace{\left(\sup_{x \in U} |||\mathrm{Id}|||_{\gamma_{F(x)} \to \sigma_{F(x)}}\right)}_{C_{\gamma \to \sigma}} \sigma_{F(x)}(u) \tag{A.3}$$

where $C_{\sigma \to \gamma} < +\infty$ and $C_{\gamma \to \sigma} < +\infty$. Clearly, $\sigma_{F(x)}$ and $\gamma_{F(x)}$ are equivalent on $\mathrm{par}(F(x))$ uniformly over $x \in U$. Therefore, there is a constant C, that can be easily expressed in terms of $C_{\sigma \to \gamma}$ and $C_{\gamma \to \sigma}$, such that for any $u \in \mathrm{par}(F(x)) \cap \mathrm{par}(F(x'))$

$$|\gamma_{F(x')}(u) - \gamma_{F(x)}(u)| \leqslant C |\sigma_{F(x')}(u) - \sigma_{F(x)}(u)| \leqslant C\beta \|u\| \|x' - x\| \ .$$

□

B. Proofs of Section 3

Proof of Proposition 1.

(i) 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. $e_x \in \bar{S}_x \cap T_x$.

(ii) This is straightforward from the fact that $S_x = \left\{\eta \in \mathbb{R}^N : \ \eta_{T_x} = 0\right\}$, $\bar{S}_x = S_x + e_x$ and $e_x \in T_x$ from (i).

□

Proof of Proposition 2. It follows from Lemma 2(v) since $0 \in \mathrm{ri}(\partial J(x) - f_x)$. □

Proof of Proposition 3. The gauge $J^x_{f_x}$ is the support function of the compact convex set

$$\mathscr{K}_x \stackrel{\text{def.}}{=} \partial J(x) - f_x = \left\{\eta \in \mathbb{R}^N : \ J^{x,\circ}_{f_x}(\eta) \leqslant 1\right\} \subset S_x \ ,$$

where the inclusion follows from Proposition 2. Observe that $0 \in \mathrm{ri}\, \mathscr{K}_x$. We then invoke Lemma 1 to get the desired claims.

□

Proof of Theorem 1. Invoking Proposition 1, we get that for every $\eta \in \partial J(x)$, $\eta_{T_x} = e_x$, and $\mathrm{P}_{T_x}(f_x) = e_x$. It remains now to uniquely characterize the part of the subdifferential lying in S_x, i.e. $\partial J(x) - e_x$. Since $f_x \in \mathrm{ri}\, \partial J(x)$, we have from the one-to-one correspondence of Lemma 2(i) and the definition of the subdifferential gauge,

$$\begin{aligned}
\eta \in \left\{\eta \in \mathbb{R}^N : \ J^{x,\circ}_{f_x}(\eta_{S_x} - \mathrm{P}_{S_x}(f_x)) \leqslant 1\right\} &\iff \eta_{S_x} - \mathrm{P}_{S_x}(f_x) \in \partial J(x) - f_x \\
&\iff \eta_{S_x} \in \partial J(x) - e_x \\
&\iff \eta \in \partial J(x) \ .
\end{aligned}$$

□

Proof of Proposition 6. 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 $(\mathscr{P}_\lambda(y))$, 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 $(\mathscr{P}_0(y))$. 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 7. Let $J = \gamma_C$, $x \in T$ and $x' \in S$.

$\Rightarrow$: We recall that $C = \{u :\ J(u) \leqslant 1\}$. By virtue of Lemma 3(iii), we have

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

Using [HUL01, Theorem V.3.3.3(iii)], we have

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

and

$$\sigma_{C\cap S}(x') = \overline{\mathrm{conv}}\left(\inf(\sigma_C(x'), \iota_T(x'))\right) = \sigma_C(x') = J^\circ(x')\ ,$$

the implication follows.

$\Leftarrow$: Using again Lemma 3, we get

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

This concludes the proof. □

Proof of Proposition 7.

Let $J=\gamma_C$. We only need to show that $J^{x,\circ}_{e_x}(\eta_{S_x}) = J^\circ(\eta_{S_x})$. This follows from Proposition 2, Lemma 7 and Lemma 3(ii). Indeed,

$$
\begin{aligned}
J^{x,\circ}_{e_x}(\eta_{S_x}) &= \inf_{\tau\geqslant 0} \max(J^\circ(\tau e_x + \eta_{S_x}), \tau) && \text{from Proposition 2,} \\
&= \inf_{\tau\geqslant 0} \max(\tau J^\circ(e_x), J^\circ(\eta_{S_x}), \tau) && \text{from Lemma 7,} \\
&= \inf_{\tau\geqslant 0} \max(J^\circ(\eta_{S_x}), \tau) && \text{from } e_x\in\partial J(x)\subset C^\circ, \\
&= J^\circ(\eta_{S_x})\ .
\end{aligned}
$$

□

C. Proofs of Section 4

Proof of Lemma 8. Let x_1, x_2 be two (global) minimizers of $(\mathscr{P}_\lambda(y))$. 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

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

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

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

LEMMA 10 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 $g : [0,1] \to \mathbb{R}$ a convex function such that $g(0) = 0$. Let $(t,s) \in (0,1]^2$ with $s > t$. Then,

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

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 $x \in C$,

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

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

We now compute the directional derivative of a finite-valued convex function J.

LEMMA 11 The directional derivative $J'(x,\delta)$ at point $x \in \mathbb{R}^N$ in the direction δ reads

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

Proof. This comes directly from the structure of $J^x_{f_x}$. Indeed, one has

$$\begin{aligned} J^x_{f_x}(\delta_{S_x}) &= J^x_{f_x}(\delta) && \text{Using Proposition 3(ii)} \\ &= \sup_{\eta \in \partial J(x) - \{f_x\}} \langle \eta, \delta \rangle \\ &= -\langle \delta, f_x \rangle + \sup_{\eta \in \partial J(x)} \langle \eta, d \rangle \\ &= -\langle \delta, f_x \rangle + J'(x,\delta) \\ &= -\langle e_x, \delta_{T_x} \rangle - \langle P_{S_x}(f_x), \delta_{S_x} \rangle + J'(x,\delta) \ . \end{aligned}$$

$\square$

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

Let x be a solution of $(\mathscr{P}_\lambda(y))$. According to Lemma 8, the set of minimizers of $(\mathscr{P}_\lambda(y))$ reads $\mathbf{M} \subseteq x + \mathrm{Ker}(\Phi)$, which is a closed convex set. We can therefore rewrite $(\mathscr{P}_\lambda(y))$ as

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

Invoking Lemma 10 with $C = \mathbf{M}$, x is thus the unique minimizer if

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

Using Lemma 11 and the fact that $\mathrm{Ker}(\Phi)$ is a subspace, this is equivalent to

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

which is ($\mathrm{NSP}^{\mathrm{S}}$). □

Proof of Corollary 1. Using [HUL01, Theorem V.2.2.3] and the fact that $J'(\cdot;\delta)$ is the support function of $\partial J(x)$, we know that

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

Applying this with $\eta = \Phi^* \alpha \in \mathrm{ri}(\partial J(x))$, and using Lemma 11, we obtain

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

Moreover, since $J^x_{f_x}$ and $\mathrm{Ker}(J^x_{f_x}) = T_x = T$ from Proposition 3(iii), and $(\mathscr{C}_T)$ holds, we get

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

We conclude using Theorem 2. □

Proof of Theorem 3.

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

$$\begin{aligned} \eta \in \left\{ \eta \in \mathbb{R}^N : \, J^{\circ}_{f_x}(\eta_S - \mathrm{P}_S(f_x)) < 1 \right\} &\iff \eta_S - \mathrm{P}_S(f_x) \in \mathrm{ri}(\partial J(x) - \{f_x\}) \\ &\iff \eta \in \mathrm{ri}(\partial J(x)) \, . \end{aligned}$$

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).

□

D. Proofs of Section 5

Proof of Theorem 4. Without loss of generality, we show this result for $\Gamma = \|\cdot\|$ since for every $x \in \mathbb{R}^N$,

$$\Gamma(x) \leqslant |||\mathrm{Id}|||_{\Gamma \to \ell^2} \|x\|.$$

Recall that J is partly smooth at x relative to T_x, and $\partial J : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ is Lipschitz-continuous around x relative to T_x.

- *Existence of f_x.* Such a mapping exists according to [AF09, Theorem 9.4.3].
- *ν-stability.* Using [Lew02, Proposition 2.10] the sharpness property Definition 7(ii) is locally stable. Hence, for $x' \in T_x$ in a neighbourhood of x, $\mathscr{T}_{T_x}(x') = T_x = T_{x'}$. The radius of this neighbourhood can be taken as ν_x.

- μ*-stability.* Using [HUL01, Corollary VI.2.1.3], we write for any $h \in T_x$

  $$J(x+th) = J(x) + t\langle s, h\rangle + o(t) = J(x) + t\langle e_x, h\rangle + o(t),$$

  where $s \in \mathbf{F}_{\partial J(x)}(h)$. Since J restricted to $T_x \cap U$ is C^2 according to the smoothness property, repeating this argument at order 2 allows to conclude that the mapping $z \in T_x \cap U \mapsto e_z$ is C^1, when local Lipschitz continuity follows immediately.

- τ*-stability.* One has

  $$J^{x,\circ}_{f_x}(\mathrm{P}_S(f_x - f_{x'})) \leqslant |||\mathrm{P}_{S_x}|||_{J^{x,\circ}_{f_x} \to \ell^2} \|f_x - f_{x'}\| \leqslant \tau_x \|x - x'\|,$$

  where $\tau_x = |||\mathrm{P}_{S_x}|||_{J^{x,\circ}_{f_x} \to \ell^2} \beta$ and β is the Lipschitz constant associated to f_x, proving (5.4).

- ξ*-stability.* By assumption, there exists a neighbourhood of x, say U, such that ∂J is κ-Lipschitz on $U \cap T_x$, and $x \mapsto f_x$ is β-Lipschitz. Hence, the mapping $x \mapsto (\partial J(x) - f_x)$ is $(\kappa + \beta)$-Lipschitz on $U \cap T_x$. Moreover, from the ν-stability, we have $S_x = \mathrm{par}(\partial(x)) = \mathrm{par}(\partial(x'))$ for all x' in $U \cap T_x$. In view of Lemma 6, we get that for any $u \in S_x$, there is a constant $C < +\infty$ such that

  $$J^{x',\circ}_{f_{x'}}(u) - J^{x,\circ}_{f_x}(u) \leqslant C(\beta + \kappa)\|x' - x\|\|u\|.$$

  Since $\|u\| \leqslant |||\mathrm{Id}|||_{\ell^2 \to J^{x,\circ}_{f_x}} J^{x,\circ}_{f_x}(u)$, we get the desired bound by setting $\xi_x = C(\beta + \kappa)|||\mathrm{Id}|||_{\ell^2 \to J^{x,\circ}_{f_x}}$.

$\square$

Proof of Proposition 8.

(i) First, we have (recall that H and G are everywhere finite)

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

Let $S^J = \mathrm{span}(\partial J(x) - \eta^J)$ and $S^G = \mathrm{span}(\partial G(x) - \eta^G)$, 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

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

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

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

(iii) As $f_x^J \in \mathrm{ri}\,\partial J(x)$ and $f_x^G \in \mathrm{ri}\,\partial G(x)$, it follows from [Roc96, Corollary 6.6.2] that

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

The subdifferential gauge associated to H is then

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

which is coercive and finite on S^H according to Proposition 2. Invoking Lemma 4, we get the desired result since for any $\rho \geqslant 0$,

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

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

□

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

It is straightforward to see that the function $\Gamma^H = \max(\Gamma^J, \Gamma^G)$ is indeed a gauge, which is finite and coercive on $T^H = T^J \cap T^G$. Moreover, given that both J and G are partly smooth relative to a subspace 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,

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

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

- Since J and G are both partly smooth relative to a subspace, then we have $T_x^J = T_{x'}^J$ and $T_x^G = T_{x'}^G$, and thus by Proposition 8(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 8(ii)

$$\begin{aligned}
\Gamma^H(e_x^H - e_{x'}^H) &= \Gamma^H\left(\mathrm{P}_{T^H}(e_x^J + e_x^G - e_{x'}^J - e_{x'}^G)\right) \\
&\leqslant \Gamma^H\left(\mathrm{P}_{T^H}(e_x^J - e_{x'}^J)\right) + \Gamma^H\left(\mathrm{P}_{T^H}(e_x^G - e_{x'}^G)\right) \\
&\leqslant |||\mathrm{P}_{T^H}|||_{\Gamma^J \to \Gamma^H} \Gamma^J\left(e_x^J - e_{x'}^J\right) + |||\mathrm{P}_{T^H}|||_{\Gamma^G \to \Gamma^H} \Gamma^G\left(e_x^G - e_{x'}^G\right) \\
&\leqslant \left(\mu_x^J |||\mathrm{P}_{T^H}|||_{\Gamma^J \to \Gamma^H} + \mu_x^G |||\mathrm{P}_{T^H}|||_{\Gamma^G \to \Gamma^H}\right) \Gamma^H(x - x') \ ,
\end{aligned}$$

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

$$\begin{aligned}
&H^{x,\circ}_{f_x^H}\left(\mathrm{P}_{S^H}(f_x^H - f_{x'}^H)\right) \\
&= H^{x,\circ}_{f_x^H}\left(\mathrm{P}_{S^J}(f_x^J - f_{x'}^J) + \mathrm{P}_{S^G}(f_x^G - f_{x'}^G) + \mathrm{P}_{S^H}(e_x^J - e_{x'}^J) + \mathrm{P}_{S^H}(e_x^G - e_{x'}^G)\right) \\
&\leqslant H^{x,\circ}_{f_x^H}\left(\mathrm{P}_{S^J}(f_x^J - f_{x'}^J)\right) + H^{x,\circ}_{f_x^H}\left(\mathrm{P}_{S^G}(f_x^G - f_{x'}^G)\right) \\
&\quad + H^{x,\circ}_{f_x^H}\left(\mathrm{P}_{S^H}(e_x^J - e_{x'}^J)\right) + H^{x,\circ}_{f_x^H}\left(\mathrm{P}_{S^H}(e_x^G - e_{x'}^G)\right) \ . \qquad \text{(A.1)}
\end{aligned}$$

According to Proposition 8(iii), we have

$$H^{x,\circ}_{f^H_x}\left(\mathrm{P}_{S^J}(f^J_x - f^J_{x'})\right) = \inf_{\eta_1+\eta_2=\mathrm{P}_{S^J}(f^J_x - f^J_{x'})} \max(J^{x,\circ}_{f^J_x}(\eta_1), G^{x,\circ}_{f^G_x}(\eta_2)) \ .$$

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

$$H^{x,\circ}_{f^H_x}\left(\mathrm{P}_{S^J}(f^J_x - f^J_{x'})\right) \leqslant J^{x,\circ}_{f^J_x}\left(\mathrm{P}_{S^J}(f^J_x - f^J_{x'})\right) .$$

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

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

That is

$$\mathrm{P}_{S^H}(e^J_x - e^J_{x'}) = \mathrm{P}_{S^H\cap T^J}(e^J_x - e^J_{x'}) \ .$$

Thus

$$H^{x,\circ}_{f^H_x}\left(\mathrm{P}_{S^H}(e^J_x - e^J_{x'})\right) \leqslant |||\mathrm{P}_{S^H\cap T^J}|||_{\Gamma^J\to H^{x,\circ}_{f^H_x}} \Gamma^J\left(e^J_x - e^J_{x'}\right) \ .$$

Similar reasoning leads to the following bounds

$$\begin{aligned}
H^{x,\circ}_{f^H_x}\left(\mathrm{P}_{S^G}(f^G_x - f^G_{x'})\right) &\leqslant G^{x,\circ}_{f^G_x}\left(\mathrm{P}_{S^G}(f^G_x - f^G_{x'})\right), \\
H^{x,\circ}_{f^H_x}\left(\mathrm{P}_{S^H}(e^G_x - e^G_{x'})\right) &\leqslant |||\mathrm{P}_{S^H\cap T^G}|||_{\Gamma^J\to H^{x,\circ}_{f^H_x}} \Gamma^G\left(e^G_x - e^G_{x'}\right) \ .
\end{aligned}$$

Having this, we can continue to bound (A.1) as

$$\begin{aligned}
&H^{x,\circ}_{f^H_x}\left(\mathrm{P}_{S^H}(f^H_x - f^H_{x'})\right) \\
&\leqslant J^{x,\circ}_{f^J_x}\left(\mathrm{P}_{S^J}(f^J_x - f^J_{x'})\right) + G^{x,\circ}_{f^G_x}\left(\mathrm{P}_{S^G}(f^G_x - f^G_{x'})\right) \\
&\quad + |||\mathrm{P}_{S^H\cap T^J}|||_{\Gamma^J\to H^{x,\circ}_{f^H_x}} \Gamma^J\left(e^J_x - e^J_{x'}\right) + |||\mathrm{P}_{S^H\cap T^G}|||_{\Gamma^J\to H^{x,\circ}_{f^H_x}} \Gamma^G\left(e^G_x - e^G_{x'}\right) \\
&\leqslant \tau^J_x \Gamma^J(x-x') + \tau^G_x \Gamma^G(x-x') + \mu^J_x |||\mathrm{P}_{S^H\cap T^J}|||_{\Gamma^J\to H^{x,\circ}_{f^H_x}} \Gamma^J\left(x-x'\right) \\
&\quad + \mu^G_x |||\mathrm{P}_{S^H\cap T^G}|||_{\Gamma^G\to H^{x,\circ}_{f^H_x}} \Gamma^G\left(x-x'\right) \\
&\leqslant \left(\tau^J_x + \tau^G_x + \mu^J_x |||\mathrm{P}_{S^H\cap T^J}|||_{\Gamma^J\to H^{x,\circ}_{f^H_x}} + \mu^G_x |||\mathrm{P}_{S^H\cap T^G}|||_{\Gamma^G\to H^{x,\circ}_{f^H_x}}\right) \Gamma^H(x-x') \ ,
\end{aligned}$$

where the last two inequalities J and G follow from μ^J_x-, τ^J_x-, μ^G_x- and τ^G_x- stability of J and G.

- ξ_x^H**-stability:** Proposition 8(iii) again yields that for any $\eta \in S^H$

$$H^\circ_{f^H_{x'}}(\eta) = \inf_{\eta_1+\eta_2=\eta} \max(J^\circ_{f^J_{x'}}(\eta_1), G^\circ_{f^G_{x'}}(\eta_2)) \\ \leqslant \max(J^\circ_{f^J_{x'}}(\bar\eta_1), G^\circ_{f^G_{x'}}(\bar\eta_2))$$

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 partly smooth relative to a subspace, hence respectively ξ_x^J- and ξ_x^G-stable. Therefore, with the form of Γ^H we have

$$J^\circ_{f^J_{x'}}(\bar\eta_1) \leqslant (1+\xi_x^J \Gamma^J(x-x')) J^{x,\circ}_{f^J_x}(\bar\eta_1) \leqslant \beta J^{x,\circ}_{f^J_x}(\bar\eta_1) \\ G^\circ_{f^G_{x'}}(\bar\eta_2) \leqslant (1+\xi_x^G \Gamma^G(x-x')) G^{x,\circ}_{f^G_x}(\bar\eta_2) \leqslant \beta G^{x,\circ}_{f^G_x}(\bar\eta_2) \; ,$$

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

$$\max(J^\circ_{f^J_{x'}}(\eta_1), G^\circ_{f^G_{x'}}(\eta_2)) \leqslant \beta \max(J^{x,\circ}_{f^J_x}(\bar\eta_1), G^{x,\circ}_{f^G_x}(\bar\eta_2)) \; .$$

Taking in particular

$$(\bar\eta_1,\bar\eta_2) \in \operatorname*{Argmin}_{\eta_1+\eta_2=\eta} \max(J^{x,\circ}_{f^J_x}(\eta_1), G^{x,\circ}_{f^G_x}(\eta_2))$$

we arrive at

$$H^\circ_{f^H_{x'}}(\eta) \leqslant \beta \inf_{\eta_1+\eta_2=\eta} \max(J^\circ_{f^J_x}(\eta_1), G^\circ_{f^G_x}(\eta_2)) = \beta H^\circ_{f^H_x}(\eta) \; .$$

This completes the proof.

□

Proof of Corollary 2. Differentiability entails that $\partial G(x) = \{\nabla G(x)\}$, whence we obtain $T_x^G = \mathbb{R}^N$ and $e_x^G = \nabla G(x)$ (see Example 3). Applying Proposition 8, we get the result. It is sufficient to remark that the smooth perturbation G translates the subdifferential $\partial J(x)$ by $\nabla G(x)$. Hence, using our choice of f_x^{J+G}, we find the same subdifferential gauge. □

Proof of Corollary 3. Since G is C^2 on $\mathbb{R}^N$, it is obviously partly smooth relative to $T_x^G = \mathbb{R}^N$ according to [Lew02, Example 3.1]. We now exhibit the constants involved.

- ν**-stability.** For every $x' \in \mathbb{R}^N$, $x' \in T_x^G$, and thus $\nu_x^G = +\infty$, implying that $\nu_x^H = \nu_x^J$.
- μ**-stability.** Using the μ-stability of J and the fact that ∇G is β-Lipschitz, we get that

$$\mu_x^H = \mu_x^J \, |||\mathrm{P}_{T^J}|||_{\Gamma^J \to \Gamma^H} + \beta \, |||\mathrm{P}_{T^J}|||_{\ell^2 \to \Gamma^H} .$$

- τ**- and** ξ**-stability.** Since $S=\{0\}$, $\tau_x^G = \xi_x^G = 0$, and we get from Proposition 9

$$\tau_x^H = \tau_x^J \quad \text{and} \quad \xi_x^H = \xi_x^J .$$

□

Proof of Proposition 10.

(i) As J is finite-valued, we have $\partial J = D \circ \partial J_0 \circ D^*$, hence $S = DS_0 = \mathrm{Im}(D_{S_0})$ and $T = S^\perp = \mathrm{Ker}(D^*_{S_0})$.

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

$$e \in \underset{z \in \bar{S}}{\operatorname{argmin}} \|z\| = \underset{z - De_0 \in S}{\operatorname{argmin}} \|z\| = De_0 + \underset{h \in S}{\operatorname{argmin}} \|h + De_0\|$$
$$= De_0 + \mathrm{P}_S(-De_0) = (\mathrm{Id} - \mathrm{P}_S) De_0 = \mathrm{P}_T De_0 = D_T e_0 \ .$$

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

$$f_{0,D^*x} \in \operatorname{ri} \partial J_0(D^*x) \Rightarrow D f_{0,D^*x} \in D \operatorname{ri} \partial J_0(D^*x)$$
$$\iff f_x \in \operatorname{ri} D \partial J_0(D^*x) \iff f_x \in \operatorname{ri} \partial J(x) \ .$$

We follow the same lines as in the proof of Lemma 5, where we additionally invoke Proposition 3(ii) to get

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

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

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

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

□

Proof of Proposition 11. In the following, all operator bounds that appear are finite owing to the coercivity assumption on the involved gauges in Definition 8 of a partly smooth regularizer.

- Let x' be such that
$$\Gamma(x - x') \leqslant \frac{1}{|||D^*|||_{\Gamma \to \Gamma_0}} \nu_{0,D^*x}.$$

Hence,

$$\Gamma_0(D^*x - D^*x') \leqslant |||D^*|||_{\Gamma\to\Gamma_0}\Gamma(x-x') \leqslant \nu_{0,D^*x}$$

As J_0 is a partly smooth relative to a subspace at D^*x, we have $T_{0,D^*x} = T_{0,D^*x'} = T_0$ and consequently, using Proposition 10(i), $T_x = \mathrm{Ker}(D^*_{S_{0,D^*x}}) = \mathrm{Ker}(D^*_{S_{0,D^*x'}}) = T_{x'} = T = S^\perp$.

- μ_x**-stability:** we now have

$$\begin{aligned}
\Gamma(e_x - e'_x) &= \Gamma(\mathrm{P}_T\, D(e_{0,D^*x} - e_{0,D^*x'})) && \text{Proposition 10(ii)}\\
&\leqslant |||D_T|||_{\Gamma_0\to\Gamma}\Gamma_0(e_{0,D^*x} - e_{0,D^*x'})\\
&\leqslant \mu_{0,D^*x}|||D_T|||_{\Gamma_0\to\Gamma}\Gamma_0(D^*x - D^*x') && \text{using } \mu_{0,D^*x}\text{-stability of } J_0\\
&\leqslant \mu_{0,D^*x}|||D_T|||_{\Gamma_0\to\Gamma}|||D^*|||_{\Gamma\to\Gamma_0}\Gamma(x-x').
\end{aligned}$$

- 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

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

Thus, subadditivity yields

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

Using Proposition 10(iii) and τ_{0,D^*x}-stability of J_0, we get the following bound on the first term

$$\begin{aligned}
&J^{x,\circ}_{f_x}(D_S\,\mathrm{P}_{S_0}(f_{0,D^*x} - f_{0,D^*x'}))\\
&= \inf_{z\in\mathrm{Ker}(D)\cap S_0} J^{D^*x,\circ}_{0,f_{0,D^*x}}(D^+_{S_0}D_S\,\mathrm{P}_{S_0}(f_{0,D^*x} - f_{0,D^*x'}) + z)\\
&\leqslant J^{D^*x,\circ}_{0,f_{0,D^*x}}(D^+_{S_0}D_S\,\mathrm{P}_{S_0}(f_{0,D^*x} - f_{0,D^*x'}))\\
&\leqslant \left|\left|\left|D^+_{S_0}D_S\right|\right|\right|_{J^{D^*x,\circ}_{0,f_{0,D^*x}}\to J^{D^*x,\circ}_{0,f_{0,D^*x}}} J^{D^*x,\circ}_{0,f_{0,D^*x}}(\mathrm{P}_{S_0}(f_{0,D^*x} - f_{0,D^*x'}))\\
&\leqslant \tau_{0,D^*x}\left|\left|\left|D^+_{S_0}D_S\right|\right|\right|_{J^{D^*x,\circ}_{0,f_{0,D^*x}}\to J^{D^*x,\circ}_{0,f_{0,D^*x}}} \Gamma_0(D^*x - D^*x')\\
&\leqslant \tau_{0,D^*x}\left|\left|\left|D^+_{S_0}D_S\right|\right|\right|_{J^{D^*x,\circ}_{0,f_{0,D^*x}}\to J^{D^*x,\circ}_{0,f_{0,D^*x}}} |||D^*|||_{\Gamma\to\Gamma_0}\Gamma(x-x').
\end{aligned}$$

Now, combining Proposition 10(iii) and μ_{0,D^*x}-stability of J_0, we obtain the following bound on the second term

$$\begin{aligned}
J^{x,\circ}_{f_x}(D_S(e_{0,D^*x} - e_{0,D^*x'})) &\leqslant J^{D^*x,\circ}_{0,f_{0,D^*x}}(D^+_{S_0}D_S(e_{0,D^*x} - e_{0,D^*x'}))\\
&\leqslant \left|\left|\left|D^+_{S_0}D_S\right|\right|\right|_{\Gamma_0\to J^{D^*x,\circ}_{0,f_{0,D^*x}}} \Gamma_0(e_{0,D^*x} - e_{0,D^*x'})\\
&\leqslant \mu_{0,D^*x}\left|\left|\left|D^+_{S_0}D_S\right|\right|\right|_{\Gamma_0\to J^{D^*x,\circ}_{0,f_{0,D^*x}}} |||D^*|||_{\Gamma\to\Gamma_0}\Gamma(x-x').
\end{aligned}$$

Combining these inequalities, we arrive at

$$J^{x,\circ}_{f_x}(\mathrm{P}_S(f_x - f_{x'})) \leqslant \Big(\tau_{0,D^*x} \left|\left|\left| D^+_{S_0} D_S \right|\right|\right|_{J^{D^*x,\circ}_{0,f_{0,D^*x}} \to J^{D^*x,\circ}_{0,f_{0,D^*x}}} + \mu_{0,D^*x} \left|\left|\left| D^+_{S_0} D_S \right|\right|\right|_{\Gamma_0 \to J^{D^*x,\circ}_{0,f_{0,D^*x}}} \Big) |||D^*|||_{\Gamma \to \Gamma_0} \Gamma(x - x'),$$

whence we get τ_x-stability.

- ξ_x**-stability:** from Proposition 10(iii), we can write for any $\eta \in S$

$$\begin{aligned} J^{x',\circ}_{f_{x'}}(\eta) &= \inf_{z \in \mathrm{Ker}(D) \cap S_0} J^{D^*x',\circ}_{0,f_{0,D^*x'}}(D^+_{S_0}\eta + z) \\ &\leqslant J^{x',\circ}_{f_{x'}}(D^+_{S_0}\eta + \bar{z}) \end{aligned}$$

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

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

$$J^{D^*x',\circ}_{0,f_{0,D^*x'}}(D^+_{S_0}\eta + \bar{z}) \leqslant \big(1 + \xi_{0,D^*x}\Gamma_0(D^*x - D^*x')\big) J^{D^*x,\circ}_{0,f_{0,D^*x}}(D^+_{S_0}\eta + \bar{z}) \ .$$

Taking in particular

$$\bar{z} \in \underset{z \in \mathrm{Ker}(D) \cap S_0}{\mathrm{Argmin}} \ J^{D^*x,\circ}_{0,f_{0,D^*x}}(D^+_{S_0}\eta + z)$$

we get the bound

$$\begin{aligned} J^{x',\circ}_{f_{x'}}(\eta) &\leqslant \big(1 + \xi_{0,D^*x}\Gamma_0(D^*x - D^*x')\big) \inf_{z \in \mathrm{Ker}(D) \cap S_0} J^{D^*x,\circ}_{0,f_{0,D^*x}}(D^+_{S_0}\eta + z) \\ &= \big(1 + \xi_{0,D^*x}\Gamma_0(D^*x - D^*x')\big) J^{x',\circ}_{f_{x'}}(\eta) \\ &= \Big(1 + \xi_{0,D^*x}|||D^*|||_{\Gamma \to \Gamma_0}\Gamma(x - x')\Big) J^{x',\circ}_{f_{x'}}(\eta) \ , \end{aligned}$$

where we used again Proposition 10(iii) in the first equality.

□

E. Proofs of Section 6

Proof of Theorem 5. 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 $e \in \mathrm{Im}(\Phi_T^*)$.

On the other hand,

$$J^{x_0,\circ}_{f_0}(\Phi_S^* \Phi_T^{+,*} e - \mathrm{P}_S f) = \mathbf{IC}(x_0) < 1.$$

□

Proof of Theorem 6. 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, by ($\mathscr{C}_T$), is the unique solution to

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

and then to show that it is actually the unique solution to $(\mathscr{P}_\lambda(y))$ under the assumptions of Theorem 6.

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

LEMMA 12 Let $x_0 \in \mathbb{R}^N$ and denote $T = T_{x_0}$. Assume that ($\mathscr{C}_T$) holds. Then $(\mathscr{P}_\lambda^T(y))$ has exactly one minimizer $\hat{x}$, and the latter satisfies

$$\hat{x} = x_0 + \Phi_T^+ w - \lambda(\Phi_T^* \Phi_T)^{-1}\tilde{e} \quad \text{where} \quad \tilde{e} \in \mathrm{P}_T(\partial J(\hat{x})). \tag{A.1}$$

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

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

Thus, using Proposition 6(i), $\hat{x}$ has to satisfy

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

for any $\tilde{e} \in \mathrm{P}_T(\partial J(\hat{x}))$. Owing to the invertibility of Φ on T, i.e. ($\mathscr{C}_T$), we obtain (A.1). □

We are now in position to prove Theorem 6. 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 $(\mathscr{P}_\lambda(y))$ using Theorem 3.

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

E.0.1 *Step 1: Subspace equality.* By construction of $\hat{x}$ in $(\mathscr{P}_\lambda^T(y))$, it is clear that $\hat{x} \in T$. The key argument now is to use that J is partly smooth relative to a subspace at x_0, and to show that

$$\Gamma(x_0 - \hat{x}) \leqslant \nu, \tag{A.2}$$

which in turn will imply subspace equality, i.e. $T_{\hat{x}} = T$ (see Definition 8).

We have from (A.1) and subadditivity that

$$\begin{aligned}
\Gamma(x_0 - \hat{x}) &\leqslant \Gamma(-\Phi_T^+ w) + \lambda \Gamma((\Phi_T^* \Phi_T)^{-1}\tilde{e}) \\
&\leqslant |||(\Phi_T^* \Phi_T)^{-1}|||_{\Gamma \to \Gamma} \{\Gamma(-\Phi_T^* w) + \lambda \Gamma(\tilde{e})\} \\
&\leqslant |||(\Phi_T^* \Phi_T)^{-1}|||_{\Gamma \to \Gamma} \{|||\Phi_T^*|||_{\ell^2 \to \Gamma}\varepsilon + \alpha_0 \lambda\} .
\end{aligned} \tag{A.3}$$

where $\alpha_0 = \Gamma(\tilde{e})$. Consequently, to show that (A.2) is verified, it is sufficient to prove that

$$A\varepsilon + B\lambda \leqslant \nu, \tag{C_1}$$

where we set the positive constants

$$A = \left|\!\left|\!\left| (\Phi_T^* \Phi_T)^{-1} \right|\!\right|\!\right|_{\Gamma \to \Gamma} |||\Phi_T^*|||_{\ell^2 \to \Gamma},$$
$$B = \alpha_0 \left|\!\left|\!\left| (\Phi_T^* \Phi_T)^{-1} \right|\!\right|\!\right|_{\Gamma \to \Gamma}.$$

Suppose for now that (C_1) holds and consequently, $T_{\hat{x}} = T$. Then decomposability of J on T (Theorem 1) implies that

$$\hat{e} = \mathrm{P}_{T_{\hat{x}}}(\partial J(\hat{x})) = \mathrm{P}_T(\partial J(\hat{x})) = \tilde{e},$$

where we have denote $\hat{e} = e_{\hat{x}}$. Thus (A.1) yields the following implicit equation

$$\hat{x} = x_0 + \Phi_T^+ w - \lambda (\Phi_T^* \Phi_T)^{-1} \hat{e}. \tag{A.4}$$

E.0.2 *Step 2: $\hat{x}$ is the unique solution of ($\mathscr{P}_\lambda(y)$).* Recall that under condition (C_1), J is decomposable at $\hat{x}$ and x_0 with the same model subspace T. Moreover, (A.4) is nothing but condition (4.1) in Theorem 3 satisfied by $\hat{x}$. To deduce that $\hat{x}$ is the unique solution of ($\mathscr{P}_\lambda(y)$), it remains to show that (4.2) holds i.e.,

$$J_{\hat{f}}^{\hat{x},\circ}(\lambda^{-1} \Phi_S^*(y - \Phi \hat{x}) - \hat{f}_S) < 1. \tag{A.5}$$

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

Under condition (C_1), the ξ-stability property (5.5) of J at x_0 yields

$$J_{\hat{f}}^{\hat{x},\circ}(\lambda^{-1} \Phi_S^*(y - \Phi \hat{x}) - \hat{f}_S) \leqslant \big(1 + \xi \Gamma(x_0 - \hat{x})\big) J_{f_0}^{x_0,\circ}(\lambda^{-1} \Phi_S^*(y - \Phi \hat{x}) - \hat{f}_S). \tag{A.6}$$

Furthermore, from (A.4), we can derive

$$\lambda^{-1} \Phi_S^*(y - \Phi \hat{x}) - \hat{f}_S = \Phi_S^* \Phi_T^{+,*} \hat{e} + \lambda^{-1} \Phi_S^* Q_T w - \hat{f}_S, \tag{A.7}$$

where $Q_T = \mathrm{Id} - \Phi_T \Phi_T^+ = \mathrm{P}_{\mathrm{Ker}(\Phi_T^*)}$. Inserting(A.7) in (A.6), we obtain

$$J_{\hat{f}}^{\hat{x},\circ}(\lambda^{-1} \Phi_S^*(y - \Phi \hat{x}) - \hat{f}_S) \leqslant \big(1 + \xi \Gamma(x_0 - \hat{x})\big) J_{f_0}^{x_0,\circ}(\Phi_S^* \Phi_T^{+,*} \hat{e} + \lambda^{-1} \Phi_S^* Q_T w - \hat{f}_S).$$

Moreover, subadditivity yields

$$\begin{aligned} J_{f_0}^{x_0,\circ}(\Phi_S^* \Phi_T^{+,*} \hat{e} + \lambda^{-1} \Phi_S^* Q_T w - \hat{f}_S) \leqslant{}& J_{f_0}^{x_0,\circ}(\Phi_S^* \Phi_T^{+,*} e - f_S) + J_{f_0}^{x_0,\circ}(\Phi_S^* \Phi_T^{+,*}(\hat{e} - e)) \\ &+ J_{f_0}^{x_0,\circ}(\mathrm{P}_S(f - \hat{f})) + J_{f_0}^{x_0,\circ}(\lambda^{-1} \Phi_S^* Q_T w). \end{aligned} \tag{A.8}$$

We now bound each term of (A.8). In the first term, one recognizes

$$J_{f_0}^{x_0,\circ}(\Phi_S^* \Phi_T^{+,*} e - f_S) \leqslant \mathbf{IC}(x_0). \tag{A.9}$$

Appealing to the μ-stability property, we get

$$\begin{aligned} J_{f_0}^{x_0,\circ}(\Phi_S^* \Phi_T^{+,*}(\hat{e} - e)) &\leqslant \left|\!\left|\!\left| -\Phi_S^* \Phi_T^{+,*} \right|\!\right|\!\right|_{\Gamma \to J_{f_0}^{x_0,\circ}} \Gamma(e - \hat{e}) \\ &\leqslant \mu \left|\!\left|\!\left| -\Phi_S^* \Phi_T^{+,*} \right|\!\right|\!\right|_{\Gamma \to J_{f_0}^{x_0,\circ}} \Gamma(x_0 - \hat{x}). \end{aligned} \tag{A.10}$$

From τ-stability, we have

$$J_{f_0}^{x_0,\circ}(f_S - \hat{f}_S) \leqslant \tau \Gamma(x_0 - \hat{x}). \tag{A.11}$$

Finally, we use a simple operator bound to get

$$J_{f_0}^{x_0,\circ}(\lambda^{-1}\Phi_S^* Q_T w) \leqslant \frac{1}{\lambda} |||\Phi_S^* Q_T|||_{\ell^2 \to J_{f_0}^{x_0,\circ}} \varepsilon. \tag{A.12}$$

Following the same steps as for the bound (A.3), except using $\tilde{e} = \hat{e}$ here, gives

$$\Gamma\big(x_0 - \hat{x}\big) \leqslant \left|\left|\left| (\Phi_T^* \Phi_T)^{-1} \right|\right|\right|_{\Gamma \to \Gamma} \{ |||\Phi_T^*|||_{\ell^2 \to \Gamma} \varepsilon + \lambda \Gamma(\hat{e}) \}. \tag{A.13}$$

Plugging inequalities (A.9)-(A.13) into (A.6) we get the upper-bound

$$\begin{aligned}
& J_{\hat{f}}^{\hat{x},\circ}(\Phi_S^* \Phi_T^{+,*} \hat{e} + \lambda^{-1} \Phi_S^* Q_T w - \hat{f}_S) \\
& \leqslant (1 + \xi \Gamma(x_0 - \hat{x})) \Big(\mathbf{IC}(x_0) + \Gamma(x_0 - \hat{x}) \big(\mu \left|\left|\left| -\Phi_S^* \Phi_T^{+,*} \right|\right|\right|_{\Gamma \to J_{f_0}^{x_0,\circ}} + \tau \big) \\
& \quad + \frac{1}{\lambda} |||\Phi_S^* Q_T|||_{\ell^2 \to J_{f_0}^{x_0,\circ}} \varepsilon \Big) \\
& \leqslant (1 + \xi (c_1 \varepsilon + \lambda c_2)) \Big(\mathbf{IC}(x_0) + (c_1 \varepsilon + \lambda c_2) \bar{\mu} + \frac{1}{\lambda} c_4 \varepsilon \Big) < 1,
\end{aligned}$$

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

$$\begin{array}{rclrcl}
c_1 & = & A, & c_2 & = & \alpha_1 \left|\left|\left| (\Phi_T^* \Phi_T)^{-1} \right|\right|\right|_{\Gamma \to \Gamma}, \\
c_3 & = & \left|\left|\left| -\Phi_S^* \Phi_T^{+,*} \right|\right|\right|_{\Gamma \to J_{f_0}^{x_0,\circ}}, & c_4 & = & \left|\left|\left| \Phi_S^* Q_T \right|\right|\right|_{\ell^2 \to J_{f_0}^{x_0,\circ}}.
\end{array}$$

If is then sufficient that

$$(1 + \xi (c_1 \varepsilon + \lambda c_2)) \Big(\mathbf{IC}(x_0) + (c_1 \varepsilon + \lambda c_2) \bar{\mu} + \frac{1}{\lambda} c_4 \varepsilon \Big) < 1, \tag{A.14}$$

for (4.2) in Theorem 3 to be in force.

In particular, if

$$C\varepsilon \leqslant \lambda$$

holds for some constant $C > 0$ to be fixed later, then inequality (A.14) is true if

$$P(\lambda) = a\lambda^2 + b\lambda + c > 0 \quad \text{where} \quad \left\{ \begin{array}{l} a = -\xi \bar{\mu} \, (c_1/C + c_2)^2 \\ b = -(c_1/C + c_2) \, (\xi \mathbf{IC}(x_0) + \xi c_4/C + \bar{\mu}) \\ c = 1 - \mathbf{IC}(x_0) - c_4/C \end{array} \right. . \tag{A.15}$$

Let us set the value of C to

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

which, for $0 \leqslant \mathbf{IC}(x_0) < 1$, 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 $\lambda_{\max}$ equal to

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

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 $\lambda_{\max}$, 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

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

imply condition (A.14), which in turn yields (A.5).

E.0.3 *Step 3:* (C_1) *and* (C_2) *are in agreement.* It remains now that show the compatibility of (C_1) and (C_2), i.e. to provide appropriate regimes of λ and ε such that both conditions hold simultaneously. We first observe that (C_1) and the left-hand-side of (C_2) both hold for λ fulfilling

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

This updates (C_2) to the following ultimate range on λ

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

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

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

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

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

which completes the proof.

$\square$

F. Proofs of Section 7

Proof of Proposition 12. The subdifferential of $\|\cdot\|_1$ reads

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

The expressions of S_x, T_x, e_x and f_x follow immediately. Since $e_x \in \operatorname{ri}\partial\|\cdot\|_1(x)$ and $\|\cdot\|_1$ is separable, it follows from Definition 6 that the ℓ^1-norm is a strong gauge. Therefore $J^\circ_{f_x} = J^\circ = \|\cdot\|_\infty$, and Proposition 7 specializes to the stated subdifferential.

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

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

This implies that $\forall i \in I(x)$, $|x'_i| > \nu_x - \|x-x'\|_\infty \geqslant 0$, 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 centred at x of radius ν_x, one can choose $\mu_x = 0$. Finally $\tau_x = \xi_x = 0$ because $f_x = e_x$. □

Proof of Proposition 14. The proof of the first part was given Section 3.1 and Section 3.2 where the ℓ^∞-norm example was considered.

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

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

This means that x' lies in the relative interior of the ℓ^1-ball (relatively to T) centred 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 $\operatorname{sign}(x_{(I(x))}) = \operatorname{sign}(x'_{(I(x'))})$. 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. □

Proof of Proposition 15. Again, the proof of the first part was given Section 3.1 and Section 3.3 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

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

ensures that for all $b \in I$

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

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 13 Given any pair of non-zero vectors u and v where, $\|u - v\| \leqslant \rho\|u\|$, for $0 < \rho < 1$, we have

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

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

$$\left\| \frac{u}{\|u\|} - \frac{v}{\|v\|} \right\|^2 = 2 - 2\frac{\langle u, v\rangle}{\|u\|\|v\|} = 2 - 2\frac{\|u\|^2 + \|u\|\|d\|\beta}{\|u\|\sqrt{\|u\|^2 + \|d\|^2 + 2\|u\|\|d\|\beta}}, \tag{A.1}$$

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

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

□

By definition of ν_x, we have $(1-\delta)\|x_b\| > \nu_x$, for $\delta \in]0,1]$, $\forall b \in I$, and thus $\|x_b - x'_b\| \leqslant \nu_x \leqslant (1-\delta)\|x_b\|$. Lemma 13 then applies, and it follows that, $\forall b \in I$

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

and therefore we get

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

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

□

Proof of Proposition 16. 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 $I = \{i \in \{1,\cdots,N_H\} :\ (x_i)_+ = J_0(x)\}$.

- If $u_i \leqslant 0$, $\forall i \in \{1,\cdots,N_H\}$, 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 $I_0 = \{i \in \{1, \cdots, N_H\} : \ u_i = J_0(u) = 0\}$. 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 $f_u \in \operatorname{ri} \partial J_0(u)$. To get the subdifferential gauge, we use some calculus rules on gauges and apply Lemma 2 to get

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

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

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

- Assume now that $u_i > 0$ for at least one $i \in \{1, \cdots, N_H\}$. In such a case, $J_0(u) = \|u\|_\infty$, and the subdifferential becomes
$$\partial J_0(u) = \Sigma_{I_+} \ ,$$
where $I_+ \{i \in \{1, \cdots, N_H\} : \ u_i = J_0(u) \quad \text{and} \quad u_i > 0\}$. The forms of S_u, T_u, e_u, f_u and the subdifferential gauge can then be retrieved from those of the ℓ^∞-norm with $s_{(I_+)} = 1$ and $s_{(I_+^c)} = 0$.

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

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

for $\delta \in]0, 1]$. This means that x' lies in the relative interior of the ℓ^1-ball (relatively to T) centred 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 $\mu_u = \tau_u = \xi_u = 0$.
□

G. Proofs of Section 8

Proof of Theorem 7. To lighten the notation, we drop the dependence on x of T, S and e. Without of loss of generality, by symmetry of the norm, we will assume that the entries of x are positive.

We follow the same program as in the compressed sensing literature, see e.g. [CR12]. The key ingredient of the proof is the fact that owing to the isotropy of the Gaussian ensemble, α_F and Φ_S^* are independent. Thus, for some $\tau > 0$

$$\Pr(\mathbf{IC}(x) \geqslant 1) \leqslant \Pr\left(\mathbf{IC}(x) \geqslant 1 \middle| \|\alpha_F\| \leqslant \tau\right) + \Pr(\|\alpha_F\| \geqslant \tau) \ .$$

As soon as $Q \geqslant \dim(T) = N - |I| + 1$, Φ_T is full-column rank. Thus

$$\|\alpha_F\|^2 = \langle e, (\Phi_T^* \Phi_T)^{-1} e \rangle \ .$$

$(\Phi_T^*\Phi_T)^{-1}$ is an inverse Wishart matrix with Q degrees of freedom. To estimate the deviation of this quadratic form, we use classical results on inverse χ^2 random variables with $Q-N+|I|$ degrees of freedom and we get the tail bound

$$\Pr\left(\|\alpha_F\| \geqslant \sqrt{\frac{1}{Q-N+|I|-t}}\|e\|\right) \leqslant e^{-\frac{t^2}{4(Q-N+|I|)}}$$

for $t>0$. Now, conditionally on α_F, the entries of $\alpha_S = \mathrm{P}_S\,\Phi^*\alpha_F$ are i.i.d. $\mathcal{N}(0,\|\alpha_F\|^2)$ and so are those of $-\alpha_S$ by trivial symmetry of the centred Gaussian. Thus, using a union bound, we get

$$\begin{aligned}
\Pr\left(\mathbf{IC}(x)\geqslant 1 \middle| \|\alpha_F\|\leqslant\tau\right) &\leqslant \Pr\left(\max_{i\in I}\,(-(\alpha_{S_x})_i)_+ \geqslant 1/|I| \middle| \|\alpha_F\|\leqslant\tau\right)\\
&\leqslant \Pr\left(\max_{i\in I}\,((\alpha_{S_x})_i)_+ \geqslant 1/|I| \middle| \|\alpha_F\|\leqslant\tau\right)\\
&\leqslant |I|\Pr\left((z)_+\geqslant 1/(\tau|I|)\right)\\
&\leqslant |I|\Pr\left(z\geqslant 1/(\tau|I|)\right)\\
&\leqslant |I|e^{-\frac{1}{2\tau^2|I|^2}}\ .
\end{aligned}$$

Observe that $(\alpha_S)_i = 0$ for all $i\in I^c$. Choosing

$$\tau = \sqrt{\frac{1}{|I|(Q-N+|I|-t)}}$$

where we used that $\|e\| = 1/\sqrt{I}$, and inserting in the above probability terms, we get

$$\begin{aligned}
\Pr\left(\|\alpha_F\|\geqslant\tau\right) &\leqslant e^{-\frac{t^2}{4(Q-N+|I|)}}\ ,\\
\Pr\left(\mathbf{IC}(x)\geqslant 1 \middle| \|\alpha_F\|\leqslant\tau\right) &\leqslant e^{-\left(\frac{Q-N+|I|-t}{2|I|}-\log(|I|/2)\right)}\ .
\end{aligned}$$

Equating the arguments of the exponentials and solving

$$\frac{t^2}{4q}+\frac{t}{2|I|}-\left(\frac{q}{2|I|}-\log\left(\frac{|I|}{2}\right)\right)=0$$

for t to get equal probabilities, we get

$$t = \frac{q}{|I|}\left(\sqrt{1+2|I|\left(1-2\frac{2|I|\log\left(\frac{|I|}{2}\right)}{q}\right)}-1\right)\ ,$$

where $q = Q-N+|I|\geqslant 1$ by the restricted injectivity assumption. Setting

$$\beta = \frac{q}{2|I|\log\left(\frac{|I|}{2}\right)}\ ,$$

we get under the bound on Q that $\beta > 1$, and

$$t = 2\beta \log\left(\frac{|I|}{2}\right)\left(\sqrt{1+2|I|\frac{\beta-1}{\beta}}-1\right) .$$

Inserting t in one of the probability terms, and after basic algebraic rearrangements, we get the probability of success with the expression of the function $f(\beta, |I|)$. ☐

References

[AF09] J.-P. Aubin and H. Frankowska. *Set-Valued Analysis*. Modern Birkhauser Classics. Birkhauser, 2009.

[ALMT13] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: A geometric theory of phase transitions in convex optimization. *arXiv preprint arXiv:1303.6672*, 2013.

[Bac08a] F. Bach. Consistency of the group lasso and multiple kernel learning. *The Journal of Machine Learning Research*, 9:1179–1225, 2008.

[Bac08b] F. Bach. Consistency of trace norm minimization. *The Journal of Machine Learning Research*, 9:1019–1048, 2008.

[Bac10] F. Bach. Structured sparsity-inducing norms through submodular functions. *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2010.

[Bak99] S. Bakin. *Adaptive regression and model selection in data mining problems*. PhD thesis, Australian National University, 1999.

[BO04] M. Burger and S. Osher. Convergence rates of convex variational regularization. *Inverse Problems*, 20(5):1411, 2004.

[Cad71] J. A. Cadzow. Algorithm for the minimum-effort problem. *IEEE Trans. Autom. Control.*, 16(1):60–63, Feb 1971.

[CDS99] 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.

[CH08] C. Chesneau and M. Hebiri. Some theoretical results on the grouped variables lasso. *Mathematical Methods of Statistics*, 17(4):317–326, 2008.

[CR12] E. Candès and B. Recht. Simple bounds for recovering low-complexity models. *Mathematical Programming*, pages 1–13, 2012.

[CRPW12] 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.

[DH01] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory*, 47(7):2845–2862, 2001.

[DT10] 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.

[EMR07] M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse problems*, 23(3):947, 2007.

[FPV$^+$13] M. J. Fadili, G. Peyré, S. Vaiter, C-A. Deledalle, and J. Salmon. Stable recovery with analysis decomposable priors. In *Proc. SampTA*, 2013.

[Fuc04] J.-J. Fuchs. On sparse representations in arbitrary redundant bases. *IEEE Transactions on Information Theory*, 50(6):1341–1344, 2004.

[Gra11] M. Grasmair. Linear convergence rates for tikhonov regularization with positively homogeneous functionals. *Inverse Problems*, 27(7):075014, 2011.

[HL04] W.L. Hare and A.S. Lewis. Identifying active constraints via partial smoothness and prox- regularity. *J. Convex Anal.*, 11(2):251–266, 2004.

[HUL01] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis And Minimization Algorithms*, volume I and II. Springer, 2001.

[JFF12] 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.

[KRZ14] M. Kabanava, H. Rauhut, and H. Zhang. Robust analysis ℓ_1-recovery from gaussian measurements and total variation minimization. Technical report, 2014.

[Lew02] A. S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization*, 13(3):702–725, 2002.

[LOS00] C. Lemaréchal, F. Oustry, and C. Sagastizábal. The *u*-lagrangian of a convex function. *Trans. Amer. Math. Soc.*, 352(2):711–729, 2000.

[LV10] Y. Lyubarskii and R. Vershynin. Uncertainty principles and vector quantization. *Information Theory, IEEE Transactions on*, 56(7):3491–3501, 2010.

[LZ09] H. Liu and J. Zhang. Estimation consistency of the group lasso and its applications. *Journal of Machine Learning Research*, 5:376–383, 2009.

[LZ13] A. S. Lewis and S. Zhang. Partial smoothness, tilt stability, and generalized hessians. *SIAM Journal on Optimization*, 23(1):74–94, 2013.

[MR11] 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.

[NDEG13] S. Nam, M.E. Davies, M. Elad, and R. Gribonval. The cosparse analysis model and algorithms. *Applied and Computational Harmonic Analysis*, 34(1):30?–56, 2013.

[NRWY10] 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.

[OB12] G. Obozinski and F. Bach. Convex relaxation for combinatorial penalties. *arXiv preprint arXiv:1205.1240*, 2012.

[OJF+12] 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.

[OTH13] S. Oymak, C. Thrampoulidis, and B. Hassibi. The squared-error of generalized LASSO: A precise analysis. Technical report, 2013.

[PT12] S. Petry and G. Tutz. Shrinkage and variable selection by polytopes. *Journal of Statistical Planning and Inference*, 142(1):48–64, 2012.

[RF08] 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.

[Roc96] R.T. Rockafellar. *Convex analysis*, volume 28. Princeton University Press, 1996.

[ROF92] L.I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[RRN12] N. Rao, B. Recht, and R. Nowak. Signal recovery in unions of subspaces with applications to compressive imaging. *arXiv preprint arXiv:1209.3079*, 2012.

[RW98] R.T. Rockafellar and R. Wets. *Variational analysis*, volume 317. Springer Verlag, 1998.

[SWB+04] 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.

[SYB12] 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.

[Tib96] R. Tibshirani. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society. Series B. Methodological*, 58(1):267–288, 1996.

[TSR+04] 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.

[VPDF13] S. Vaiter, G. Peyré, C. Dossal, and M. J. Fadili. Robust sparse analysis regularization. *IEEE Transactions on Information Theory*, 59(4):2001–2016, 2013.

[VPF13] S. Vaiter, G. Peyré, and M. J. Fadili. Robust polyhedral regularization. In *Proc. SampTA*, 2013.

[VPF14] S. Vaiter, G. Peyré, and J. Fadili. Model consistency of partly smooth regularizers. Technical report, Preprint Hal-00987293, 2014.

[WH10] F. Wei and J. Huang. Consistent group selection in high-dimensional linear regression. *Bernoulli*, 16(4):1369–1384, 2010.

[Wri93] S. J. Wright. Identifiable surfaces in constrained optimization. *SIAM Jounal Control Optimisation*, 31(4):1063–1079, 1993.

[YL05] 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.

[ZH05] H. Zou and T. Hastie. Regularization and variable selection via the elastic net. *J. Royal Statistical Soc. B*, 67(2):301–320, 2005.