Sharp, strong, and unique minimizers for low complexity robust recovery

Jalal Fadili[*] Tran T. A. Nghia[†] Trinh T. T. Tran[‡]

Abstract

In this paper, we show the important roles of sharp minima and strong minima for robust recovery. We also obtain several characterizations of sharp minima for convex regularized optimization problems. Our characterizations are quantitative and verifiable especially for the case of decomposable norm regularized problems including sparsity, group-sparsity, and low-rank convex problems. For group-sparsity optimization problems, we show that a unique solution is a strong solution and obtain quantitative characterizations for solution uniqueness.

Contents

[*]Normandie Université, ENSICAEN, UNICAEN, CNRS, GREYC, France; email: Jalal.Fadili@greyc.ensicaen.fr

[†]Department of Mathematics and Statistics, Oakland University, Rochester, MI 48309, USA; email: nttran@oakland.edu. Research of this author was supported by the US National Science Foundation under grant DMS-1816386.

[‡]Department of Mathematics and Statistics, Oakland University, Rochester, MI 48309, USA; email: thitutrinhtran@oakland.edu

1 Introduction

1.1 Problem statement

Inverse problems and regularization theory is a central theme in various areas of engineering and science. A typical case is where one observes a vector $y_0 \in \mathbb{R}^m$ of linear measurements according to

$$y_0 = \Phi x_0 \tag{1.1}$$

where $\Phi \in \mathbb{R}^{m\times n}$ is known, and is typically an idealization of the sensing mechanism in signal/imaging science applications, or the design matrix in a parametric statistical regression problem. The vector $x_0 \in \mathbb{R}^n$ is the unknown quantity of interest.

Solving the inverse problem associated to the linear forward model (1.1) amounts to recovering x_0, either exactly or to a good approximation, knowing y_0 and Φ. This is however a quite challenging task especially when the linear system (1.1) is underdetermined with $m \ll n$. In fact, even when $m = n$, Φ is in general ill-conditioned or even singular. This entails that the linear inverse problem is in general ill-posed.

In order to bring back the problem to the land of well-posedness, it is necessary to restrict the inversion process to a well-chosen subset of $\mathbb{R}^n$ containing the plausible solutions including x_0. This can be achieved by adopting a variational framework and solving the following optimization problem

$$\min_{x\in\mathbb{R}^n} \quad J(x) \quad \text{subject to} \quad \Phi x = y_0, \tag{1.2}$$

where J is function which is bounded from below (wlog non-negative). The function J, known as a regularizer, is designed in such way that it is the smallest on the sought-after solutions. Popular example in signal/image processing and machine learning is the ℓ_1 norm to promote sparsity, the ℓ_1/ℓ_2 norm to promote group sparsity, analysis sparsity seminorm (i.e., $J = J_0 \circ D^*$, $D^* : \mathbb{R}^n \to \mathbb{R}^p$ is linear and J_0 is the ℓ_1 or ℓ_1/ℓ_2 norm); see [6, 10, 13, 24, 32, 36, 38, 44, 46] to cite a few. Another popular example is the nuclear norm, i.e., the ℓ_1 norm of the singular values of a matrix, to recover low-rank matrices [11, 9]. These prototypical cases are included in the class of decomposable norms [29, 8, 15] that covers a wide range of optimization-based recovery problems in different areas of data science.

When the observation is subject to errors, the system (1.1) is modified to

$$y = \Phi x_0 + \omega \tag{1.3}$$

where ω accounts either for noise and/or modeling errors. The errors can be either deterministic (in this case, one typically assumes to know some bound on $\|\omega\|$) or random (in which case its distribution is assumed to be known). Throughout, ω will be assumed deterministic with $\|\omega\| \leq \delta$, with $\delta > 0$ known.

To recover x_0 from (1.3), one solves the noise-aware form (aka Mozorov regularization in the inverse problems literature)

$$\min_{x \in \mathbb{R}^n} J(x) \quad \text{subject to} \quad \|y - \Phi x\| \le \delta \tag{1.4}$$

or its penalized/Lagrangian form (aka Tikhonov regularization in the inverse problems literature)

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|\Phi x - y\|^2 + \mu J(x). \tag{1.5}$$

with regularization parameter $\mu > 0$. It is well-known that (1.4) and (1.5) are formally equivalent in the sense that there exists a bijection between δ and μ such that the both problems share the same set of solutions.

We say that *robust recovery* occurs when solutions to (1.4) or (1.5) are close enough to the original signal x_0 as long as μ is appropriately chosen (as a function of the noise level δ). For ℓ_1-regularized optimization problem, a significant result was achieved in [20, Theorem 4.7], which shows that when μ is proportional to δ, the error between any solution of problem (1.5) to x_0 is $\mathcal{O}(\delta)$[1] if and only if x_0 is the unique solution of (1.2). The structure of the ℓ_1 norm plays a crucial role in their result. Sufficient (but not necessary in general) conditions for robust recovery with a linear convergence rate have been proposed for general class of regularizers by solving either (1.4) or (1.5); see [11, 15, 42, 45, 39] and references therein. For instance [11] and [39] provide a sufficient condition via the geometric notion of *descent cone* for robust recovery with linear rate by solving (1.4) (but not (1.5)). Injectivity of Φ on the descent cone turns out to be a sufficient and necessary condition for solution uniqueness of (1.2). However, solution uniqueness is not enough to guarantee robust recovery with linear rate for general regularizers.

In fact, we observe that [11] needs x_0 to be a *sharp* solution of problem (1.2). A sharp minimum, introduced by Crome and Polyak [12, 30, 31], is a point around which the objective value has a first-order growth (see Section 2 for a rigorous definition and discussion). It is certainly sufficient for solution uniqueness and indeed is called *strong uniqueness* in [12]. In the case of polyhedral problems at which the function $J(x)$ in (1.2) is piecewise affine, e.g., problem (1.2) with ℓ_1 or ℓ_∞ norms, solution uniqueness is equivalent to sharp minima. This is the key reason why [20] only needs solution uniqueness for robust recovery with linear rate. In the optimization community, sharp minima is well-known for its essence to finite convergence of the proximal point algorithms; see, e.g., [31, 34].

Sharp minima is actually hidden in both *exact recovery* (by solving (1.2)) and robust recovery. As we will show later, sharpness is closely related to many other results on exact and robust recovery, for instance those involving conditions based on non-degenerate dual certificates (aka *Nondegenerate Source Condition*) and restricted injectivity of Φ, or the *Null Space Property* and its variants; see [13, 10, 40, 17, 41, 9, 6, 20, 16, 44, 15, 8, 44, 42, 45, 48]. When Φ is drawn from an appropriate random ensemble, existing sample complexiy bounds (i.e., a lower-bound on m depending on the "intrinsic" dimension of the model subspace containing x_0 and logarithmically on n) rely on verifying the afore-mentioned conditions with high probability on Φ; see [16, 45, 39] for comprehensive reviews. For Gaussian measurements, sample complexity bounds for exact recovery are provided in terms of the Gaussian width in [11] and the statistical dimension in [1], and the latter has been shown to precise predictions about the quantitative aspects of the phase transition for exact recovery from Gaussian measurements.

Another popular notion in optimization [5, 35] is that of *strong minima*, which is a sufficient condition for solution uniqueness. It is also used implicitly in many results on robust recovery.

[1]This justifies the conventional terminology "linear convergence rate".

A strong minimum is a point around which the objective value has a quadratic (or second-order) growth. Strong minima are certainly weaker than sharp minima. In the ℓ_1 case, we observe that the cost function in (1.5) belongs to the bigger class of convex *piecewise linear-quadratic functions* [35]. A unique minimizer of these functions is also a strong minimizer. This simple observation was used recently in [4] to obtain some characterizations for solution uniqueness to (1.5) for the case of the Lasso problem. Necessary and sufficient conditions for solution uniqueness to (1.5) where J is piecewise linear are also studied in [18, 27, 48].

1.2 Contributions and relation to prior work

The chief goal of this paper is to show the intricate and important roles of sharp minima and strong minima in robust recovery. Our main contributions are as follows (see also the diagram in Figure 1).

- As discussed above, solution uniqueness and sharp minima play a pivotal role for robust recovery with linear rate. However, except the case where the regularizer J is merely polyhedral, a unique solution may not be sharp. For example, as we will show in our Section 5, a unique solution of the group-sparsity problem (1.2) is always a strong minimum but not a sharp one. The natural question is therefore whether strong minima could also lead to some meaningful results for robust recovery. Section 3 answers this question positively without even needing convexity of J, unlike the previous work of [42, 45] which showed that a sharp minimizer implicitly guarantees robust recovery with linear rate, but convexity was essential there. We show that if x_0 is a strong solution of (1.2), then we have robust recovery by solving (1.4) or (1.5) with a convergence rate $\mathcal{O}(\sqrt{\delta})$ (Theorem 3.12). This rate can be improved to the linear rate $\mathcal{O}(\delta)$ when $J = \|\cdot\|$. Whether this linear rate for robust recovery can be achieved for other regularizers under the sole strong minimality assumption is an open question. Besides that main result, we also revisit in Theorem 3.10 robust recovery results when x_0 is assumed to be a sharp minimizer by proving linear convergence and providing more explicit bounds in comparison to the results in [11, 15, 20, 48].

- Our second main contribution, at the heart of Section 4, is to study necessary and sufficient conditions for the properties of sharp and strong minima. As mentioned above, solution uniqueness is characterized via the descent cone [11]. But it is hard to compute the descent cone for general functions J. On the other hand, sharp and strong minima can be characterized explicitly via first and second-order analysis [5, 31, 35]. We leverage these tools (see Theorem 4.6 and Proposition 4.13) to get equivalent charaterizations of unique/sharp/strong solutions, and to obtain a quantitative condition for convex regularized problems that allows us to check numerically that a solution is sharp. This condition bears similarities with the Nondegenerate Source Condition of [20, 42, 45]. Our approach does not rely on any polyhedral structure. This distinguishes our results from several ones in the literature as in [18, 27, 48], and allows us to work on more general optimization problems. We also believe that non-smooth second-order analysis of the type we develop here, which is a powerful tool in variational analysis and nonsmooth optimization, is not well-known in the robust recovery literature, and are thus of particular importance in this context.

- Finally, Section 5 is devoted to a comprehensive treatment of unique/strong solutions for group-sparsity minimization problems [32, 36, 46], an important class of (1.2) where J is the ℓ_1/ℓ_2 norm. Sufficient conditions for solution uniqueness for group-sparsity minimization problems were achieved in [8, 19, 24, 32, 36, 42, 45], but none of them are necessary. To find a characterization for solution uniqueness, we first obtain a closed form for the descent cone (Theorem 5.1). To the best of our knowledge, this was an open question in the literature, though the *statistical dimension* of this cone is more computable [1]. By relying again on second-order analysis [5, 35], we show that a unique solution of group-sparsity minimization problems (1.2) and (1.5) is indeed a strong solution (Theorem 5.3). This result is interesting in two ways: first, the function in (1.2) is not piecewise linear-quadratic, which rules out the approach developed in [4] to unify solution uniqueness and strong minima; second, strong solutions are possibly more natural than sharp solutions for non-polyhedral minimization problems whenever solution uniqueness occurs. Moreover, we establish a quantitative condition for checking unique/strong solutions to group-sparsity problems that is equivalent to solving a smooth convex optimization problem. This convex problem can be solved via available packages such as `cvxopt` as we will illustrate in the numerical results of Section 6. Consequently, we show that solution uniqueness to group-sparsity problem is equivalent to the robust recovery with rate $\mathcal{O}(\sqrt{\delta})$ for both problems (1.4) and (1.5).

Remark 1.1. *Although we do not have a rigorous study for the case where J is the nuclear norm in this paper, we provide several examples showing that a unique solution of* (1.2) *with the nuclear norm is neither sharp nor strong. This raises the challenge of understanding the interplay between unique, sharp, and strong solutions in this case. Some open questions will be discussed in Section 7 for further research in this direction.*

2 Preliminaries

Throughout the paper, $\mathbb{R}^n$ is the Euclidean space with dimension n. In $\mathbb{R}^n$ we denote the inner product by $\langle \cdot, \cdot \rangle$, the Euclidean norm by $\|\cdot\|$, and the closed ball of radius $r > 0$ centered at $x \in \mathbb{R}^n$ by $\mathbb{B}_r(x)$, and $\mathbb{I}$ the identity operator.

2.1 Some linear algebra

Given an $m \times n$ matrix A, $\operatorname{Ker} A$ (resp. $\operatorname{Im} A$) is the kernel/null (resp. image) space of A. The Moore-Penrose generalized inverse [25, Page 423] of A is denoted by $A^\dagger$, which satisfies the property

$$AA^\dagger A = A. \tag{2.1}$$

Moreover, we have

$$\begin{aligned} Ax = b \text{ is consistent if and only if } AA^\dagger b = b, \text{ and} \\ \text{the set of solutions to } Ax = b \text{ is } A^\dagger b + \operatorname{Ker} A. \end{aligned} \tag{2.2}$$

If A is injective then $A^\dagger = (A^\top A)^{-1} A^\top$, where $A^\top$ is the transpose of A. If A is surjective then $A^\dagger = A^\top (AA\top)^{-1}$. A^* will also denote the adjoint of A as a linear operator.

Suppose that $\bar{x}$ is a solution of the linear system $Ax = b$. The projection of $x \in \mathbb{R}^n$ onto the affine set $C \stackrel{\text{def}}{=} A^{-1}b$ is

$$\mathrm{P}_C(x) = (\mathbb{I} - A^\dagger A)(x - \bar{x}) + \bar{x} = x - A^\dagger A(x - \bar{x}); \tag{2.3}$$

see, e.g., [25, page 435–437]. Let $\|\cdot\|_{\mathcal{A}}$ and $\|\cdot\|_{\mathcal{B}}$ be two arbitrary norms in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively. The matrix operator norm $\|\cdot\|_{\mathcal{A},\mathcal{B}} : \mathbb{R}^{m\times n} \to \mathbb{R}_+$ is defined by

$$\|A\|_{\mathcal{A},\mathcal{B}} \stackrel{\text{def}}{=} \max_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_{\mathcal{A}}}{\|x\|_{\mathcal{B}}}.$$

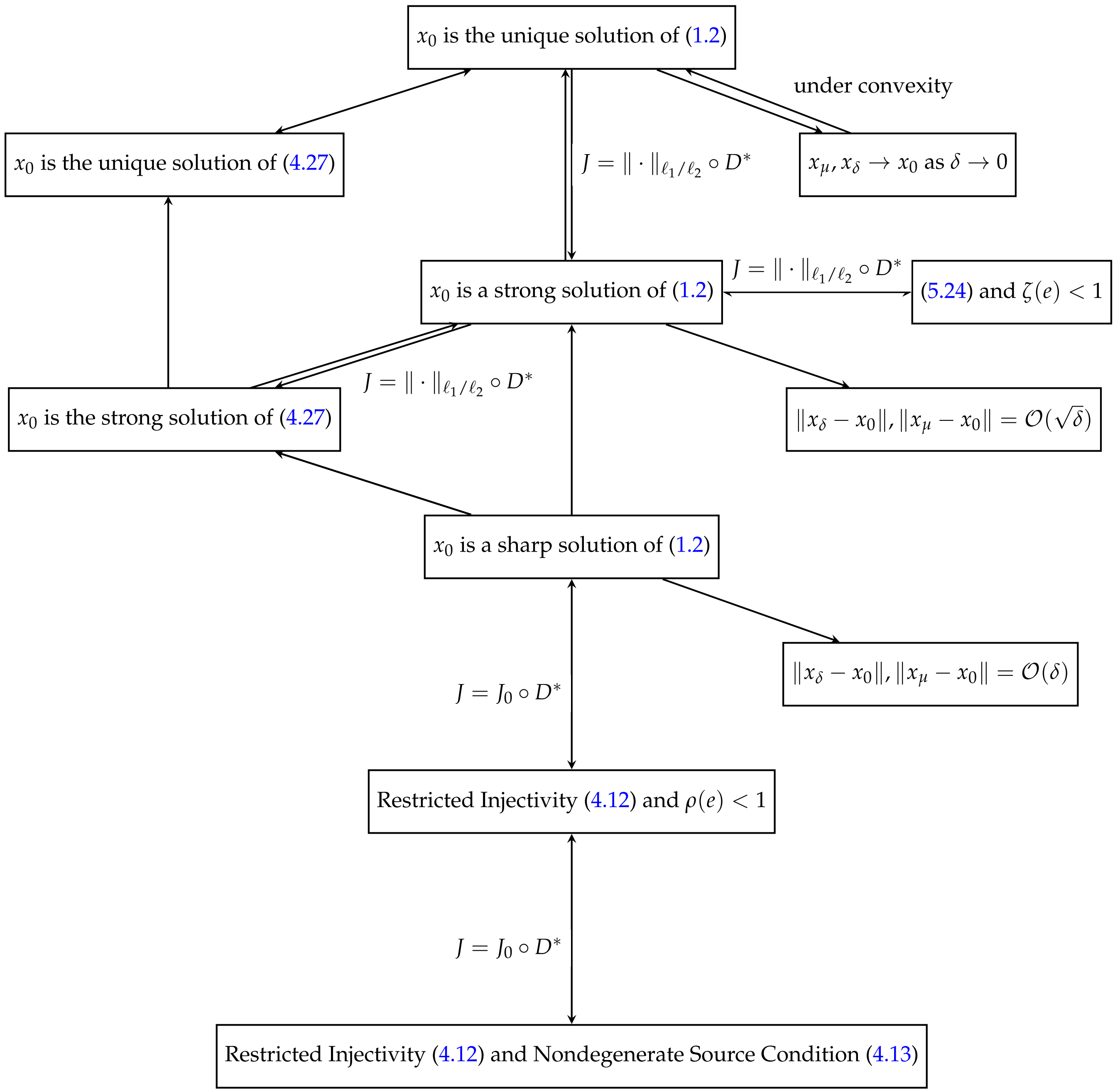


Figure 1: Diagram summarizing our main results, and in particular, the relationships between the different optimization problems and the notions of sharp/strong/unique solution. Each bidirectional arrow indicates equivalence, while simple arrows indicate the implication direction.

We write $\|A\|_{2,2}$ for $\|A\|$. The Frobenius norm of A is known as

$$\|A\|_F = \sqrt{\mathrm{Trace}\,(A^*A)}.$$

Recall that the nuclear norm of A is

$$\|A\|_* = \sum_{k=1}^{\min(m,n)} \sigma_k,$$

where σ_k, $1 \le k \le \min(m,n)$ are all the singular values of A.

2.2 Some convex analysis

Let $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}} \stackrel{\text{def}}{=} \mathbb{R} \cup \{+\infty\}$ be a proper lower semi-continuous (lsc) convex function with domain $\mathrm{dom}\,\varphi \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n |\ \varphi(x) < \infty\}$. The subdifferential of φ at $x \in \mathrm{dom}\,\varphi$ is defined by

$$\partial\varphi(x) \stackrel{\text{def}}{=} \{v \in \mathbb{R}^n |\ \varphi(u) \ge \varphi(x) + \langle v,\, u - x\rangle\,, \quad \forall u \in \mathbb{R}^n\}\,.$$

When $\varphi(\cdot) = \iota_C(\cdot)$, the indicator function of a nonempty closed convex set C, i.e., $\iota_C(x) = 0$ if $x \in C$ and $+\infty$ otherwise, the subdifferential of ι_C at $x \in C$ is the *normal cone* to C at x:

$$N_C(x) \stackrel{\text{def}}{=} \{v \in \mathbb{R}^n |\ \langle v,\, u - x\rangle \le 0, \quad \forall u \in C\}\,. \tag{2.4}$$

For a set $\Omega \subset \mathbb{R}^n$, $\mathrm{cone}\,\Omega$ is its *conical hull*, $\mathrm{int}\,\Omega$ is the interior of Ω, and $\mathrm{ri}\,\Omega$ is the relative interior of the convex set Ω. Given a convex cone $K \subset \mathbb{R}^n$, the *polar* of K is defined to be the cone

$$K^\circ \stackrel{\text{def}}{=} \{v \in \mathbb{R}^n |\ \langle v,\, u\rangle \le 0, \quad \forall u \in \Omega\}\,. \tag{2.5}$$

It is well known, see [35, Corollary 6.21] that K° is closed and convex, and that

$$K^\circ = \mathrm{cl}\,(K)^\circ \ \text{ and } \ K^{\circ\circ} = (K^\circ)^\circ = \mathrm{cl}\,(K), \tag{2.6}$$

where cl denotes the topological *closure*. We will also use cl for the closure of a function, i.e., the topological closure of its epigraph.

The *gauge function* of Ω is defined by

$$\gamma_\Omega(u) \stackrel{\text{def}}{=} \inf\{r > 0 |\ u \in r\Omega\}\,, \quad \text{for} \quad u \in \mathbb{R}^n. \tag{2.7}$$

The *support function* to Ω is denoted by

$$\sigma_\Omega(v) \stackrel{\text{def}}{=} \sup_{x\in\Omega} \langle v,\, x\rangle\,, \quad \text{for} \quad v \in \mathbb{R}^n. \tag{2.8}$$

It is a proper lsc convex function as soon as Ω is non-empty. The *polar set* and *polar cone* of Ω are given by

$$\Omega^\circ \stackrel{\text{def}}{=} \{v \in \mathbb{R}^n |\ \sigma_C(v) \le 1\} \qquad \text{and} \qquad \Omega^* \stackrel{\text{def}}{=} \{v \in \mathbb{R}^n |\ \sigma_C(v) \le 0\}\,, \tag{2.9}$$

respectively. When Ω is a non-empty convex set, Ω° is a non-empty closed convex set containing the origin. When Ω is also closed and contains the origin, it is well-known that

$$\gamma_\Omega = \sigma_{\Omega^\circ}, \tag{2.10}$$

see, e.g., [33, Corollary 15.1.2]. In this case, γ_C is a non-negative lsc convex and positively homogenous function.

2.3 Sharp and strong minima

Let us first recall the definition of sharp and strong solutions in [5, 12, 31, 35] that are playing central roles throughout the paper.

Definition 2.1 (Sharp and strong minima)**.** *We say $\bar{x}$ to be a sharp solution/minimizer of the (non necessarily convex) function φ with a constant $c > 0$ if there exists $\varepsilon > 0$ such that*

$$\varphi(x) \geq \varphi(\bar{x}) + c\|x - \bar{x}\|, \quad \forall x \in \mathbb{B}_\varepsilon(\bar{x}). \tag{2.11}$$

Moreover, $\bar{x}$ is said to be a strong solution/minimizer with a constant $\kappa > 0$ if there exists $\delta > 0$ such that

$$\varphi(x) \geq \varphi(\bar{x}) + \frac{\kappa}{2}\|x - \bar{x}\|^2, \quad \forall x \in \mathbb{B}_\delta(\bar{x}). \tag{2.12}$$

Properties (2.11) and (2.12) are also known as respectively, first order and second order (or quadratic) growth properties.

Strong minima is certainly weaker than sharp minima. Moreover, if $\bar{x}$ is a sharp or strong solution of φ, it is a unique local minimizer of φ (and the unique minimizer if φ is convex). When the function φ is a *piecewise linear convex* function, i.e., its epigraph is a polyhedron, $\bar{x}$ is a unique solution of φ if and only if it is a sharp solution. Furthermore, if φ is a convex *piecewise linear-quadratic* function in the sense that its domain is a union of finitely many polyhedral sets, relative to each of which φ is a quadratic function, $\bar{x}$ is a strong solution if and only if it is a unique solution, see, e.g., [4].

To characterize sharp and strong solutions, it is typical to use directional derivative and second subderivative; see, e.g., [5, 35].

Definition 2.2 (directional derivative and second subderivative)**.** *Let $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper lsc convex function. The directional derivative of φ at $\bar{x} \in \operatorname{dom} \varphi$ is the function $d\varphi(\bar{x}) : \mathbb{R}^n \to \overline{\mathbb{R}}$ defined by*

$$d\varphi(\bar{x})(w) \stackrel{\text{def}}{=} \lim_{t \downarrow 0} \frac{\varphi(\bar{x} + tw) - \varphi(\bar{x})}{t} \quad \text{for} \quad w \in \mathbb{R}^n. \tag{2.13}$$

The second subderivative of φ at $\bar{x} \in \operatorname{dom} \varphi$ for $\bar{v} \in \partial\varphi(\bar{x})$ is the function $d^2\varphi(\bar{x}|\bar{v}) : \mathbb{R}^n \to \overline{\mathbb{R}}_+$ defined by

$$d^2\varphi(\bar{x}|\bar{v})(w) \stackrel{\text{def}}{=} \liminf_{t \downarrow 0, w' \to w} \frac{\varphi(\bar{x} + tw') - \varphi(\bar{x}) - t\langle \bar{v}, w'\rangle}{\frac{1}{2}t^2} \quad \text{for} \quad w \in \mathbb{R}^n. \tag{2.14}$$

It is well-known that $d\varphi(\bar{x})(w) = \max\{\langle v, w\rangle \mid v \in \partial\varphi(\bar{x})\}$ if φ is continuous at $\bar{x}$. The calculation of $d^2\varphi(\bar{x}|\bar{v})(w)$ is quite involved in general; see, e.g., [5, 35, 26]. Note that from [35, Proposition 13.5 and Proposition 13.20], the function $d^2\varphi(\bar{x}|\bar{v})$ is non-negative, lsc, positively homogeneous of degree 2, and

$$\operatorname{dom} d^2\varphi(\bar{x}|\bar{v}) \subset \mathcal{T}(\bar{x}|\bar{v}) \stackrel{\text{def}}{=} \{w \in \mathbb{R}^n \mid d\varphi(\bar{x})(w) = \langle \bar{v}, w\rangle\}. \tag{2.15}$$

If, moreover, φ is twice-differentiable at $\bar{x}$, we have (see [35, Example 13.8])

$$d^2\varphi(\bar{x}|\nabla\varphi(\bar{x}))(w) = \langle w, \nabla^2\varphi(\bar{x})w\rangle, \tag{2.16}$$

where $\nabla^2\varphi(\bar{x})$ stands for the second-order derivative of the function φ at $\bar{x}$. The following sum rule for second subderivatives will be useful.

Lemma 2.3. *Let $\varphi, \phi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper lsc convex functions with $\bar{x} \in \operatorname{dom} \varphi \cap \operatorname{dom} \phi$. Suppose that φ is twice-differentiable at $\bar{x} \in \operatorname{int}(\operatorname{dom} \varphi)$ and $\bar{v} \in \partial(\varphi + \phi)(\bar{x}) = \nabla \varphi(\bar{x}) + \partial \phi(\bar{x})$. Then we have*

$$d^2(\varphi + \phi)(\bar{x}|\bar{v})(w) = \langle w, \nabla^2 \varphi(\bar{x}) w \rangle + d^2 \phi(\bar{x}|\bar{v} - \nabla \varphi(\bar{x}))(w) \quad \text{for all} \quad w \in \mathbb{R}^n. \tag{2.17}$$

Proof. Use convexity of φ, ϕ and twice-differentiability of φ at $\bar{x}$ into [35, Proposition 13.19]. □

The next result taken from [31, Lemma 3, Chapter 5] and [35, Theorem 13.24] characterize sharp and strong minima.

Lemma 2.4 (Characterization of sharp and strong minima)**.** *Let $\varphi : \mathbb{R}^n \to \overline{\mathbb{R}}$ be a proper lsc convex function with $\bar{x} \in \operatorname{dom} \varphi$. We have:*

(i) *$\bar{x}$ is a sharp minimizer of φ if and only if there exists $c > 0$ such that $d\varphi(\bar{x})(w) \geq c\|w\|$ for all $w \in \mathbb{R}^n$, i.e., $d\varphi(\bar{x})(w) > 0$ for all $w \in \mathbb{R}^n \setminus \{0\}$.*

(ii) *$\bar{x}$ is a strong minimizer to φ if and only if $0 \in \partial \varphi(\bar{x})$ and there exists $\kappa > 0$ such that $d^2 \varphi(\bar{x}|0)(w) \geq \kappa \|w\|^2$ for all $w \in \mathbb{R}^n$, which is equivalent to*

$$\operatorname{Ker} d^2 \varphi(\bar{x}|0) \stackrel{\text{def}}{=} \{w \in \mathbb{R}^n |\ d^2 \varphi(\bar{x}|0)(w) = 0\} = \{0\}.$$

It is important to observe that although the sharpness of a minimizer as defined in (2.11) is a local property, it is actually a global one for convex problems.

Lemma 2.5 (Global sharp minima)**.** *Suppose that $\bar{x}$ is a sharp solution of the proper lsc convex function φ. Then there exists $c > 0$ such that*

$$\varphi(x) - \varphi(\bar{x}) \geq c\|x - \bar{x}\| \qquad \text{for all} \quad x \in \mathbb{R}^n.$$

Proof. Since $\bar{x}$ is a sharp solution of φ, we have from Lemma 2.4(i) that $\exists c > 0$ such that $d\varphi(\bar{x})(w) \geq c\|w\|$, $\forall w \in \mathbb{R}^n$. Thus, by the characterization of the directional derivative for convex functions, we have for any $x \in \mathbb{R}^n$

$$\varphi(x) - \varphi(\bar{x}) = \varphi(\bar{x} + (x - \bar{x})) - \varphi(\bar{x}) \geq \inf_{t>0} \frac{\varphi(\bar{x} + t(x - \bar{x})) - \varphi(\bar{x})}{t} = d\varphi(\bar{x})(x - \bar{x}) \geq c\|x - \bar{x}\|.$$

The proof is complete. □

3 Unique, sharp, and strong solutions for robust recovery

3.1 Uniqueness and robust recovery

Let us consider the optimization problem in (1.2) where we suppose that $J : \mathbb{R}^n \to \overline{\mathbb{R}}$ is a non-negative lsc function but not necessarily convex. We denote by $J_\infty : \mathbb{R}^n \to \overline{\mathbb{R}}$ the *asymptotic (or horizon) function* [2] associated with J, which is defined by

$$J_\infty(w) \stackrel{\text{def}}{=} \liminf_{w' \to w, t \to \infty} \frac{J(tw')}{t}. \tag{3.1}$$

Throughout this section, we assume that J satisfies

$$\operatorname{Ker} J_\infty \cap \operatorname{Ker} \Phi = \{0\}, \tag{3.2}$$

where $\operatorname{Ker} J_\infty \stackrel{\text{def}}{=} \{w \in \mathbb{R}^n |\ J_\infty(w) = 0\}$, and the range of J_∞ is on $\mathbb{R}_+$ since J is non-negative. It then follows from [2, Corollary 3.1.2] that (3.2) ensures that problem (1.2) has a non-empty compact set of minimizers.

Define $\Psi : \mathbb{R}^n \to \overline{\mathbb{R}}$ as

$$\Psi(x) \stackrel{\text{def}}{=} J(x) + \iota_{\Phi^{-1}(\Phi x_0)}(x) \quad \text{for all} \quad x \in \mathbb{R}^n. \tag{3.3}$$

We say that x_0 is a unique, sharp, or strong solution of problem (1.2) if it is a unique, sharp, or strong solution of the minimization of the function Ψ, respectively.

Our first result shows that solution uniqueness is sufficient for robust recovery. If J is also convex, then uniqueness is also necessary. The sufficiency part of our result generalizes the corresponding part of [21, Theorem 3.5] to nonconvex problems.

Proposition 3.1 (Solution uniqueness for robust recovery)**.** *Suppose that J is a non-negative lsc function which satisfies* (3.2)*.*

(i) *If x_0 is the unique solution of problem* (1.2) *then:*

  (a) *any solution x_δ to problem* (1.4) *with $\|y - y_0\| \le \delta$ converges to x_0 as $\delta \to 0$,*

  (b) *for any constant $c_1 > 0$, any solution x_μ to problem* (1.5) *with $\mu = c_1\delta$ and $\|y - y_0\| \le \delta$ converges to x_0 as $\delta \to 0$.*

(ii) *Conversely, if J is convex and* $\operatorname{dom} J = \mathbb{R}^n$*, then:*

  (a) (i)-(a) *implies x_0 is the unique solution of* (1.2)*.*

  (b) (i)-(b) *implies x_0 is the unique solution of* (1.2)*.*

Proof. (i) Suppose that x_0 is the unique solution of problem (1.2). To justify (i)-(a), define the function $g \stackrel{\text{def}}{=} \iota_{\mathbb{B}_\delta(y)} \circ \Phi$. We have, using [2, Propositions 2.6.1, 2.6.3 and 2.1.2], that

$$\begin{aligned}(J+g)_\infty(w) \ge J_\infty(w) + g_\infty(w) = J_\infty(w) + (\iota_{\mathbb{B}_\delta(y)})_\infty(\Phi w) &= J_\infty(w) + \iota_{\{0\}}(\Phi w)\\ &= J_\infty(w) + \iota_{\operatorname{Ker}\Phi}(w) \ge 0.\end{aligned}$$

Thus, under (3.2), the set of solutions to (1.4) is nonempty and compact thanks to [2, Corollary 3.1.2]. If $\{x_\delta\}_{\delta>0}$ is unbounded, without passing to subsequences suppose that $\|x_\delta\| \to \infty$ and $\dfrac{x_\delta}{\|x_\delta\|} \to w$ as $\delta \downarrow 0$ with $\|w\| = 1$. We have

$$0 \le \left\| \Phi\left(\frac{x_\delta}{\|x_\delta\|}\right) - \frac{y}{\|x_\delta\|} \right\| \le \frac{\delta}{\|x_\delta\|},$$

and after passing to the limit, we get that $w \in \operatorname{Ker}\Phi$. Moreover, since x_0 is a feasible point of (1.4), we have $J(x_\delta) \le J(x_0)$, and thus

$$0 = \liminf_{\delta\downarrow 0} \frac{J(x_0)}{\|x_\delta\|} \ge \liminf_{\delta\downarrow 0} \frac{J(x_\delta)}{\|x_\delta\|} \ge J_\infty(w) \ge 0,$$

which means that $w \in \operatorname{Ker} J_\infty$. This entails that $0 \ne w \in \operatorname{Ker} J_\infty \cap \operatorname{Ker}\Phi$, which contradicts (3.2). Thus $\{x_\delta\}_{\delta>0}$ is bounded. Pick any subsequence $\{x_{\delta_k}\}_{k\in\mathbb{N}}$ of $\{x_\delta\}_{\delta>0}$ converging to some $\bar{x}$ as $\delta_k \downarrow 0$. From lower semicontinuity of J and of the norm, we get

$$J(\bar{x}) \le \liminf_k J(x_{\delta_k}) \le J(x_0) \ \text{ and } \ 0 \le \|y_0 - \Phi\bar{x}\| \le \liminf_k \|y_0 - \Phi x_{\delta_k}\| \le 2\lim_k \delta_k = 0.$$

This means that $\bar{x}$ is a solution of (1.2), and by uniqueness of the minimizer, $\bar{x} = x_0$. Since the subsequence $\{x_{\delta_k}\}_{k\in\mathbb{N}}$ was arbitrary, we have $x_\delta \to x_0$ as $\delta \downarrow 0$, whence claim (i)-(a) follows.

The proof of (i)-(b) uses a similar reasoning. Let $g = \frac{1}{2}\|y - \cdot\|^2 \circ \Phi$. One has, using [2, Propositions 2.6.1, 2.6.3 and 2.6.7], that

$$\begin{aligned}(\mu J + g)_\infty(w) \geq \mu J_\infty(w) + g_\infty(w) = \mu J_\infty(w) + \frac{1}{2}\left(\|\cdot\|^2\right)_\infty(\Phi w) &= \mu J_\infty(w) + \iota_{\{0\}}(\Phi w)\\ &= \mu J_\infty(w) + \iota_{\operatorname{Ker}\Phi}(w) \geq 0.\end{aligned}$$

Hence the set of optimal solutions to problem (1.5) is nonempty and compact according to [2, Corollary 3.1.2]. Similarly to part (i)-(a), we claim that $\{x_\mu\}_{\mu>0}$ is bounded. By contradiction, suppose again that $\|x_\mu\| \to \infty$ and $\frac{x_\mu}{\|x_\mu\|} \to z$ with $\|z\| = 1$. Note that by optimality of x_μ

$$\mu J(x_\mu) \leq \frac{1}{2}\|\Phi x_\mu - y\|^2 + \mu J(x_\mu) \leq \frac{1}{2}\|\Phi x_0 - y\|^2 + \mu J(x_0) \leq \frac{\delta^2}{2} + \mu J(x_0). \tag{3.4}$$

As $\mu = c_1\delta$, we obtain from (3.4) that

$$0 = \liminf_{\delta\downarrow 0} \frac{J(x_0) + \frac{\delta}{2c_1}}{\|x_\mu\|} \geq \liminf_{\delta\downarrow 0} \frac{J(x_\mu)}{\|x_\mu\|} \geq J_\infty(z) \geq 0,$$

which means $z \in \operatorname{Ker} J_\infty$. Furthermore, (3.4) also entails

$$0 \leq \frac{1}{2}\left\|\Phi\left(\frac{x_\mu}{\|x_\mu\|}\right) - \frac{y}{\|x_\mu\|}\right\|^2 \leq \frac{1}{\|x_\mu\|^2}\left(\frac{\delta^2}{2} + \mu J(x_0)\right),$$

which tells us, after passing to the limit, that $0 \neq z \in \operatorname{Ker} J_\infty \cap \operatorname{Ker}\Phi$, hence contradicting (3.2). Thus $\{x_\mu\}_{\mu>0}$ is bounded. Arguing as in (i)-(a), we use lower semicontinuity of J and of the norm to show that any cluster point of $\{x_\mu\}_{\mu>0}$ is a solution of (1.2), and deduce claim (i)-(b) thanks to uniqueness of the minimizer x_0.

(ii) Suppose now that J is also convex and $\operatorname{dom} J = \mathbb{R}^n$ (hence continuous). If (i)-(a) holds, we have

$$J(x_\delta) \leq J(x) \quad \text{for any} \quad x \quad \text{such that} \quad \|y - \Phi x\| \leq \delta \ \text{ and } \ \|y - y_0\| \leq \delta. \tag{3.5}$$

It follows that

$$J(x_\delta) \leq J(x) \quad \text{for any} \quad x \quad \text{such that} \quad \Phi x = y_0.$$

By letting $\delta \downarrow 0$, and using continuity of J, we obtain that x_0 is a minimizer of (1.2). Let $\bar{x}$ be an arbitrary minimizer of problem (1.2). As $\operatorname{dom} J = \mathbb{R}^n$, Fermat's rule for (1.2) and subdifferential calculus gives

$$0 \in \partial\Psi(\bar{x}) = \partial J(\bar{x}) + N_{\Phi^{-1}y_0}(\bar{x}) = \partial J(\bar{x}) + \operatorname{Im}\Phi^*. \tag{3.6}$$

Or, equivalently, there exists a dual multiplier $\eta \in \mathbb{R}^m$ such that $\Phi^*\eta \in \partial J(\bar{x})$[2].

If $\eta = 0$, we have $0 \in \partial J(\bar{x})$, which means that $\bar{x}$ is a minimizer to J. Thus, $\bar{x}$ is also a minimizer to problem (1.4) for any $\delta > 0$. By (i)-(a), $\bar{x} - x_0 \to 0$ as $\delta \downarrow 0$, i.e., $\bar{x} = x_0$.

[2]This is known as the Source Condition [37].

If $\eta \neq 0$, take any $\delta > 0$ and define $\mu \stackrel{\text{def}}{=} \frac{\delta}{\|\eta\|}$ and $y \stackrel{\text{def}}{=} y_0 + \mu\eta = \Phi\bar{x} + \mu\eta$. It follows that $\|y - \Phi\bar{x}\| = \|y - y_0\| = \delta$ and that

$$\frac{1}{\mu}\Phi^*(y - \Phi\bar{x}) = \Phi^*\eta \in \partial J(\bar{x}). \tag{3.7}$$

On the other hand, under our assumption, we have that $x^\star$ is a solution of (1.4) if and only if

$$0 \in \partial J(x^\star) + \Phi^* N_{\mathbb{B}_\delta(y)}(\Phi x^\star),$$

where $N_{\mathbb{B}_\delta(y)}(\Phi x^\star) = \mathbb{R}_+(\Phi x^\star - y)$ whenever $\|y - \Phi x^\star\| = \delta$. This is precisely the optimality condition verified by $\bar{x}$ in (3.7), which shows that $\bar{x}$ is also a solution of (1.4). By (i)-(a) again, we have $\bar{x} = x_0$. Since the choice of $\bar{x}$ is arbitrary, x_0 is the unique solution of (1.2).

Finally, suppose that (i)-(b) is satisfied. We have for any $x \in \Phi^{-1}y_0$ that

$$\mu J(x_\mu) \leq \frac{1}{2}\|\Phi x - y\|^2 + \mu J(x) \leq \frac{\delta^2}{2} + \mu J(x).$$

Dividing both sides by $\mu = c_1\delta$ and letting $\delta \to 0$ entails $J(x_0) \leq J(x)$ for any $x \in \Phi^{-1}y_0$, which is equivalent to saying that x_0 is a solution of problem (1.2). The proof of uniqueness follows similar lines as for problem (1.4) with the choice $y = \Phi\bar{x} + \mu\eta$ and $\mu = c_1\delta$. Note that under our assumption, Fermat's rule for problem (1.5) is (3.7) without further qualification condition unlike what is required for (1.2).

□

Remark 3.2. *In the convex case, the condition* (3.2) *is superfluous in Proposition 3.1(i). Indeed,* (3.2) *is equivalent in this case to the fact the set of minimizers of* (1.2) *is a non-empty compact set, which is obviously the case when x_0 is the unique minimizer.*

Our assumption $\operatorname{dom} J = \mathbb{R}^n$ *in part (ii) of Proposition 3.1, though verified in typical convex regularizers can be relaxed. We only need a domain qualification condition for the subdifferential calculus sum rule* (3.6) *to apply (and passing to the limit as $\delta \downarrow 0$ remains valid under mere lower semicontinuity of J). The sum rule of convex subdifferential calculus holds under various conditions, for instance if* $\operatorname{ri}(\operatorname{dom} J) \cap \Phi^{-1}(y_0) \neq \emptyset$*; see, e.g., [3, Theorem 16.37].*

3.2 Sharp minima and robust recovery

A sufficient and necessary condition for uniqueness to problem (1.2), together with its implication for robust recovery, is studied in [11] via the key geometric notion of *descent cone* that we recall now.

Definition 3.3 (Descent cone)**.** *The descent cone of the function J at x_0 is defined by*

$$\mathcal{D}_J(x_0) \stackrel{\text{def}}{=} \operatorname{cone}\left\{x - x_0 |\ J(x) \leq J(x_0)\right\}. \tag{3.8}$$

The following proposition provides a characterization for solution uniqueness whose proof can be found in [11, Proposition 2.1].

Proposition 3.4 (Descent cone for solution uniqueness [1, 11])**.** *Then x_0 is the unique solution of* (1.2) *if and only if* $\operatorname{Ker}\Phi \cap \mathcal{D}_J(x_0) = \{0\}$.

In [11, Proposition 2.2] (see also [39, Proposition 2.6]), the authors also prove a robust recovery result for solutions of (1.4) via the descent cone.

Proposition 3.5 (Robust recovery via descent cone [11])**.** *Let Suppose that there exists some $\alpha > 0$ such that $\|\Phi w\| \geq \alpha \|w\|$ for all $w \in \mathcal{D}_J(x_0)$. Then any solution x_δ to problem (1.4) satisfies*

$$\|x_\delta - x_0\| \leq \frac{2\delta}{\alpha}. \tag{3.9}$$

The parameter α is known as the minimum conic singular value of Φ with respect to the cone $\mathcal{D}_J(x_0)$. One may observe that the condition in Proposition 3.5 is not the same with the one for solution uniqueness in Proposition 3.4. The reason is that the descent cone $\mathcal{D}_J(x_0)$ is not necessarily closed in general. To see this, we first unveil the relation between the descent and *critical* cones. Denote

$$L_{x_0}(J) \stackrel{\text{def}}{=} \{x \in \mathbb{R}^n |\ J(x) \leq J(x_0)\}$$

the *sublevel set* of J at x_0. The *critical cone* of a convex function J at x_0 is

$$\mathcal{C}_J(x_0) \stackrel{\text{def}}{=} \{w \in \mathbb{R}^n |\ dJ(x_0)(w) \leq 0\}. \tag{3.10}$$

Lemma 3.6 (Relationship between descent cone and critical cone)**.** *Suppose that J is a continuous convex function. Then,*

$$\mathcal{D}_J(x_0) \subset \mathrm{cl}\,(\mathcal{D}_J(x_0)) \subset \mathcal{C}_J(x_0). \tag{3.11}$$

If moreover $0 \notin \partial J(x_0)$, then

$$\mathrm{cl}\,(\mathcal{D}_J(x_0)) = \mathcal{C}_J(x_0). \tag{3.12}$$

Proof. We have from (2.6) that

$$\mathrm{cl}\,(\mathcal{D}_J(x_0)) = (\mathcal{D}_J(x_0))^{**} = ((\mathrm{cl}\,\mathcal{D}_J(x_0))^*)^* = (N_{L_{x_0}(J)}(x_0))^* \subset \mathcal{C}_J(x_0), \tag{3.13}$$

where we used (2.6) in the first two equalities, polarity in the third, and convexity and the first claim of [35, Proposition 10.3] in the last equality. If in addition $0 \notin \partial J(x_0)$, then using [33, Corollary 23.7.1] and the last claim of [35, Proposition 10.3], the inclusion in (3.13) becomes

$$\mathrm{cl}\,(\mathcal{D}_J(x_0)) = (N_{L_{x_0}(J)}(x_0))^* = (\mathrm{cone}\,\partial J(x_0))^* = \mathcal{C}_J(x_0). \tag{3.14}$$

□

When J is also positively homogenous, the requirement $0 \notin \partial J(x_0)$ in Lemma 3.6 can be dropped.

Proposition 3.7. *Suppose that $J = \sigma_C$, where $C \subset \mathbb{R}^n$ is a non-empty compact convex set with $0 \in \mathrm{ri}\, C$. Then, (3.12) holds at any $x_0 \in \mathbb{R}^n$.*

Proof. Thanks to the assumptions on C, we have from [33, Theorem 13.2 and Corollary 13.3.1] that J convex, positively and homogenous and finite everywhere, hence continuous. We now only consider the case where $0 \in \partial J(x_0)$ as, otherwise, the result claim follows from Lemma 3.6. By [43, Lemma 1], J is also non-negative and $\mathrm{Ker}\, J$ is a linear subspace. This entails that $0 \in \partial J(x_0)$ is equivalent to $x_0 \in \mathrm{Ker}\, J$, and thus from (3.8), that

$$\mathcal{D}_J(x_0) = \mathrm{cone}\,(\mathrm{Ker}\, J - x_0) = \mathrm{Ker}\, J.$$

On the other hand, we have from (3.10) that

$$\mathcal{C}_J(x_0) = (\partial J(x_0))^* = (\operatorname{cone} \partial J(x_0))^* = N_{\partial J(x_0)}(0).$$

Moreover, by [35, Corollary 8.25]

$$\partial J(x_0) = C \cap \{x_0\}^\perp.$$

Combining this with [35, Theorem 6.42] (recall that C is convex and $0 \in \operatorname{ri} C \cap \{x_0\}^\perp$), we get

$$\mathcal{C}_J(x_0) = N_{C\cap\{x_0\}^\perp}(0) = N_C(0) + \mathbb{R}x_0 = \operatorname{Ker} J + \mathbb{R}x_0 = \operatorname{Ker} J,$$

where the last equality comes from the fact that $\operatorname{Ker} J$ is a linear subspace containing x_0. This completes the proof. □

We now turn to showing that the condition in Proposition 3.5 actually means that x_0 is a sharp solution of (1.2).

Proposition 3.8 (Characterization of sharp minima via the descent and critical cones)**.** *Let J be a continuous convex function.*

(i) *x_0 is a sharp solution of* (1.2) *if and only if* $\operatorname{Ker} \Phi \cap \mathcal{C}_J(x_0) = \{0\}$.

(ii) *If $J = \sigma_C$, where $C \subset \mathbb{R}^n$ is a non-empty compact convex set with $0 \in \operatorname{ri} C$, then x_0 is a sharp solution of* (1.2) *if and only if there exists $\alpha > 0$ such that $\|\Phi w\| \geq \alpha\|w\|$ for all $w \in \mathcal{D}_J(x_0)$.*

Proof. (i) Recall Ψ from (3.3). We have

$$d\Psi(x_0)(w) = dJ(x_0)(w) + \iota_{\operatorname{Ker}\Phi}(w). \tag{3.15}$$

By Lemma 2.4, x_0 is a sharp solution of (1.2) if and only if $d\Psi(x_0)(w) > 0$ for all $w \in \mathbb{R}^n \setminus \{0\}$. Combining this with (3.10) and (3.15) leads the claimed equivalence.

(ii) Assume x_0 is a sharp solution of (1.2). Since $\mathcal{C}_J(x_0)$ is closed, we have from claim (i) that there exists $\alpha > 0$ such that $\|\Phi w\| \geq \alpha\|w\|$ for all $w \in \mathcal{C}_J(x_0)$. We then get the first implication from the inclusion (3.11). Conversely, if there is $\alpha > 0$ such that $\|\Phi w\| \geq \alpha\|w\|$ for all $w \in \mathcal{D}_J(x_0)$, we have $\|\Phi w\| \geq \alpha\|w\|$ for all $w \in \mathcal{C}_J(x_0)$ thanks to Proposition 3.7. It follows that $\operatorname{Ker} \Phi \cap \mathcal{C}_J(x_0) = \{0\}$. This implies, in view of claim (i), that x_0 is a sharp solution of (1.2).

□

The main lesson we learn from the above, in particular Proposition 3.4 and Proposition 3.8, is that there is a gap between solution uniqueness and solution sharpness (hence robust recovery via Proposition 3.5) for problem (1.2). This difference lies in the closedness of the descent cone. Indeed, when the descent cone $\mathcal{D}_J(x_0)$ is closed, it coincides with the critical cone $\mathcal{C}_J(x_0)$. Unfortunately, in general, $\mathcal{D}_J(x_0)$ is not closed. A prominent example where $\mathcal{D}_J(x_0)$ may fail to be closed is that of the ℓ_1/ℓ_2 norm very popular to promote group sparsity. The following example is a prelude of our precise formula for the descent cone of the ℓ_1/ℓ_2 norm in Theorem 5.1, in which the interior of the critical cone $\mathcal{C}_J(x_0)$ plays a significant role.

Example 3.9 (Gap between solution uniqueness and sharpness for group-sparsity)**.** Consider the following ℓ_1/ℓ_2 norm minimization problem:

$$\min_{x\in\mathbb{R}^3} J(x) = \sqrt{x_1^2 + x_2^2} + |x_3| \quad \text{subject to} \quad \Phi x = y_0, \tag{3.16}$$

with $\Phi = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$, $x_0 = (0,1,0)^\top$, and $y_0 = \Phi x_0 = (1,0)^\top$. We have $\operatorname{Ker}\Phi = \mathbb{R}(1,-1,1)^\top$, and feasible points of (3.16) then take the form $x = x_0 + t(1,-1,1)^\top$ for any $t \in \mathbb{R}$. For all such points, we have

$$J(x) - J(x_0) = \sqrt{t^2 + (1-t)^2} + |t| - 1 = \frac{t^2 + 2(|t| - t)}{\sqrt{t^2 + (1-t)^2} + 1 - |t|},$$

which tells us that x_0 is the unique solution of (3.16). In fact, simple calculation shows that x_0 is a strong but not a sharp solution of (3.16). For J in (3.16), Figure 2 illustrates the difference between the descent cone computed by Theorem 5.1 and the critical cone (3.10) whose computation is straightforward in this case. It is obvious that $\operatorname{Ker}\Phi \cap \mathcal{C}_J(x_0) \neq \{0\}$, though $\operatorname{Ker}\Phi \cap \mathcal{D}_J(x_0) = \{0\}$ (which is indeed equivalent to solution uniqueness). □

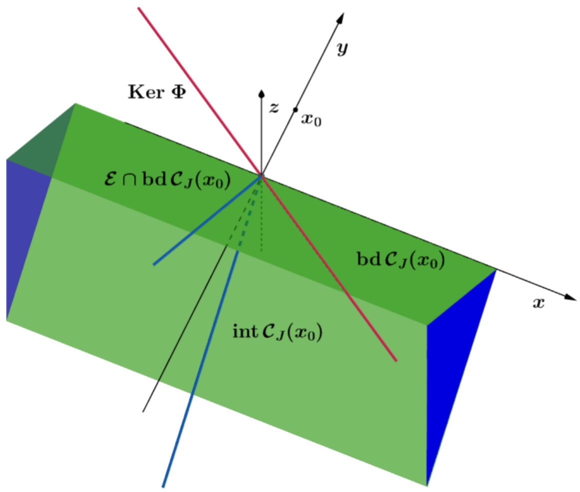


Figure 2: The descent cone $\mathcal{D}_J(x_0)$ and critical cone $\mathcal{C}_J(x_0)$ for J in (3.16). Here, $\mathcal{D}_J(x_0) = (\mathcal{E} \cap \operatorname{bd}\mathcal{C}_J(x_0)) \cup \operatorname{int}\mathcal{C}_J(x_0)$, where $\mathcal{E} = \{(0, x_2, x_3) | \, x_2, x_3 \in \mathbb{R}\}$ (see (5.8)) and $\mathcal{C}_J(x_0) = \{(x_1, x_2, x_3) | \, x_2 + |x_3| \leq 0\}$. The green shaded surface is $\operatorname{bd}\mathcal{C}_J(x_0)$ and the blue half-lines correspond to $\mathcal{E} \cap \operatorname{bd}\mathcal{C}_J(x_0)$.

Inspired by Proposition 3.5, we show next that sharpness of the minimizer guarantees robust recovery for both problems (1.4) and (1.5) with linear rate. Unlike [11, Proposition 2.2] and many other results [15, 20, 42, 45, 48] in this direction, we do not need convexity of J. In fact, the key is to assume that x_0 is a unique and sharp solution of problem (1.2).

Theorem 3.10 (Sharp minima for robust recovery)**.** *Suppose that J is a non-negative lsc function which satisfies (3.2), and is Lipschitz continuous around x_0 with modulus L. If x_0 is the unique and sharp solution of (1.2) with constant $c > 0$, then for all $\delta > 0$ sufficiently small, the following statements hold:*

(i) *Any solution x_δ to problem (1.4) satisfies*

$$\|x_\delta - x_0\| \leq \frac{2(L+c)\|\Phi^\dagger\|}{c}\delta. \tag{3.17}$$

(ii) *For any $c_1 > 0$ and $\mu = c_1\delta$, any minimizer x_μ to (1.5) satisfies*

$$\|x_\mu - x_0\| \leq \frac{c_1}{2c}\left(\frac{1}{c_1} + (L+c)\|\Phi^\dagger\|\right)^2 \delta. \tag{3.18}$$

Proof. (i) L-Lipschitz continuity of J around x_0 entails that $\exists \varepsilon > 0$ such that

$$|J(x) - J(z)| \leq L\|x - z\| \quad \text{for all} \quad x, z \in \mathbb{B}_\varepsilon(x_0). \tag{3.19}$$

Since x_0 is a sharp solution of (1.2) with constant c, there exists $\eta > 0$ such that

$$J(x) - J(x_0) \geq c\|x - x_0\| \quad \text{for all} \quad x \in \Phi^{-1}y_0 \cap \mathbb{B}_\eta(x_0). \tag{3.20}$$

Let $\bar{x}_\delta$ be the projection of x_δ onto $\Phi^{-1}y_0$. Since $x_0 \in \Phi^{-1}y_0$, we have

$$\|x_\delta - \bar{x}_\delta\| = \min_{x \in \Phi^{-1}y_0} \|x_\delta - x\| \leq \|x_\delta - x_0\|.$$

In turn

$$\|\bar{x}_\delta - x_0\| \leq \|x_\delta - \bar{x}_\delta\| + \|x_\delta - x_0\| \leq 2\|x_\delta - x_0\|, \tag{3.21}$$

Since x_0 is the unique solution of problem (1.2), Proposition 3.1(i) tells us that for all δ small enough, $x_\delta \in \mathbb{B}_{\min\{\varepsilon,\eta\}/2}(x_0)$ and, by (3.21), $\bar{x}_\delta \in \mathbb{B}_{\min\{\varepsilon,\eta\}}(x_0)$, i.e., $\bar{x}_\delta, x_\delta \in \mathbb{B}_{\min\{\varepsilon,\eta\}}(x_0)$ with ε, η in (3.19) and (3.20). We then have

$$\begin{aligned}
L\|\bar{x}_\delta - x_\delta\| &\geq J(\bar{x}_\delta) - J(x_\delta) && \text{by (3.19)},\\
&\geq J(\bar{x}_\delta) - J(x_0) && \text{by optimality of } x_0,\\
&\geq c\|\bar{x}_\delta - x_0\| && \text{by (3.20)},\\
&\geq c\|x_\delta - x_0\| - c\|x_\delta - \bar{x}_\delta\|.
\end{aligned}$$

Combining this with the projection formula (2.3) tells us that

$$\begin{aligned}
c\|x_\delta - x_0\| &\leq (L + c)\|x_\delta - \bar{x}_\delta\| = (L + c)\|\Phi^\dagger\Phi(x_\delta - x_0)\|\\
&\leq (L + c)\|\Phi^\dagger\|\|\Phi(x_\delta - x_0)\|\\
&\leq (L + c)\|\Phi^\dagger\|(\|y - \Phi x_\delta\| + \|y - \Phi x_0\|)\\
&\leq (L + c)\|\Phi^\dagger\|2\delta,
\end{aligned}$$

which is (3.17).

(ii) Optimality of x_μ gives

$$\begin{aligned}
J(x_\mu) - J(x_0) &\leq \frac{1}{2\mu}\left(\|y - \Phi x_0\|^2 - \|y - \Phi x_\mu\|^2\right)\\
&\leq \frac{1}{2\mu}\left(\|y - \Phi x_0\|^2 - \left(\|y - \Phi x_0\| - \|\Phi(x_\mu - x_0)\|\right)^2\right)\\
&\leq \frac{1}{2\mu}\left(2\delta\|\Phi(x_\mu - x_0)\| - \|\Phi(x_\mu - x_0)\|^2\right).
\end{aligned} \tag{3.22}$$

Let $\bar{x}_\mu$ be the projection of x_μ to $\Phi^{-1}y_0$. Inequality (3.21) remains valid for x_μ and $\bar{x}_\mu$ replacing x_δ and $\bar{x}_\delta$ respectively. Moreover, arguing as in the proof of claim (i), but now invoking Proposition 3.1(ii), we infer that for δ sufficiently small, $x_\mu, \bar{x}_\mu \in \mathbb{B}_{\min\{\varepsilon,\eta\}}(\bar{x})$, where ε, η are those in (3.19) and (3.20). Denote for short $\alpha = \|\Phi(x_\mu - x_0)\|$. We then get

$$\begin{aligned}
J(x_\mu) - J(x_0) &= J(\bar{x}_\mu) - J(x_0) + J(x_\mu) - J(\bar{x}_\mu)\\
&\geq c\|\bar{x}_\mu - x_0\| - L\|x_\mu - \bar{x}_\mu\| && \text{by (3.20) and (3.19)},\\
&\geq c\|x_\mu - x_0\| - (L + c)\|x_\mu - \bar{x}_\mu\|\\
&= c\|x_\mu - x_0\| - (L + c)\|\Phi^\dagger\Phi(x_\mu - x_0)\| && \text{by (2.3)},\\
&\geq c\|x_\mu - x_0\| - (L + c)\|\Phi^\dagger\|\alpha.
\end{aligned}$$

This together with (3.22) gives us that

$$
\begin{aligned}
c\|x_\mu - x_0\| &\le \frac{2\delta\alpha - \alpha^2}{2\mu} + (L+c)\|\Phi^\dagger\|\alpha \\
&= \left(\frac{\delta}{\mu} + (L+c)\|\Phi^\dagger\|\right)\alpha - \frac{\alpha^2}{2\mu} \\
&\le \frac{\mu}{2}\left(\frac{\delta}{\mu} + (L+c)\|\Phi^\dagger\|\right)^2,
\end{aligned}
\tag{3.23}
$$

which is clearly (3.18) for the choice $\mu = c_1\delta$.

□

The bounds (3.17) and (3.18) in Theorem 3.10 hold without convexity with the proviso that δ is sufficiently small. However, when the function J is convex and continuous, the bounds are satisfied for any $\delta > 0$. It is also worth pointing out that the upper bound in (3.18) is greater than or equal to the one in (3.17) due the Cauchy-Schwarz inequality. The equality occurs when we choose a special value for μ with $c_1 = \dfrac{1}{(L+c)\|\Phi^\dagger\|}$.

Corollary 3.11 (Sharp minima for robust recovery under convexity). *Suppose that J is a convex continuous function. If x_0 is a sharp solution of* (1.2)*, then statements* (i) *and* (ii) *of Theorem 3.10 hold for any $\delta > 0$ with some constant $L > 0$.*

Proof. When J is a continuous convex function and x_0 is a sharp solution of (1.2), it follows from Lemma 2.5 that the sharpness property (3.20) is global and x_0 is also the unique solution of (1.2). In turn, J satisfies (3.2) in view of Remark 3.2. Therefore, the nets $\{x_\delta\}_{\delta>0}, \{x_\mu\}_{\mu>0}$ are bounded as proved in Proposition 3.1. Thus, there exists some $R > 0$ such that $x_0, x_\delta, \bar{x}_\delta, x_\mu, \bar{x}_\mu \in \mathbb{B}_R(0)$. Convexity and continuity of J also imply that it is Lipschitz continuous on $\mathbb{B}_{2R}(x_0)$ with some Lipschitz constant $L > 0$. Overall, this tells us that we can take ε, η in (3.19) and (3.20) in the proof of Theorem 3.10 by $2R$. The rest of the proof remains valid, whence our claim follows. □

Discussion of related work. It is worth noting that Corollary 3.11 covers many results in [15, 20, 42, 14, 45, 48]. When $J = \|\cdot\|_{\mathcal{A}}$ is a norm in $\mathbb{R}^n$, claim (i) of Corollary 3.11 is exactly [11, Proposition 2.2] (see Proposition 3.5) thanks to Proposition 3.8(ii). In this case the Lipschitz constant L of J is $\|\mathbb{I}\|_{\mathcal{A},2}$.

For the case $J = \|\cdot\|_{\ell_1}$, Corollary 3.11 returns [20, Theorem 4.7] (see also [14, Proposition 1]) whose proof is less transparent and involves deriving linear convergence rate of the Bregman divergence of J, together with the characterizations of solution uniqueness to ℓ_1 optimization problems via the so-called restricted injectivity and non-degenerate source condition.

When $J(x) = \|D^*x\|_{\mathcal{A}}$ with D being an $n \times p$ matrix and $\|\cdot\|_{\mathcal{A}}$ being a norm in $\mathbb{R}^p$, the Lipschitz constant L of J is $\|\mathbb{I}\|_{\mathcal{A},2}\|D^*\|$. Corollary 3.11 covers [15, Theorem 2]. Another result in this direction is [48, Theorem 2], which only obtains linear rate for $\|D^*(x_\mu - x_0)\|$ with extra nontrivial assumptions on D.

When J is a general convex continuous regularizer as in Corollary 3.11, [42, 45] use the so-called restricted injectivity and non-degenerate source condition at x_0 to obtain robust recovery. In our forthcoming Theorem 4.6, we will show that these two conditions are equivalent to x_0 being a sharp solution[3]. It means that Corollary 3.11 is equivalent to [42, Theorem 6.1] or [45, Theorem 2].

[3] The authors in [42, 45] have already proved that these two conditions are sufficient for x_0 to be a sharp solution.

However, our path to proving robust recovery is radically different. On the one hand, [42, Theorem 6.1] generalizes the proof strategy initiated in [20] and use decomposability of J and other arguments that heavily rely on convexity of J. On the other hand, Corollary 3.11 provides a direct proof that involve natural geometrical properties of J around x_0. Most importantly, Theorem 3.10 suggests that robust recovery with linear rate occurs without convexity and reveals the crucial role played by sharpness of the minimizer to achieve robust recovery with linear rate. Comparing the constants in our bounds (3.18)-(3.17) and those in [42, 45], those in [42, 45] depend for instance on a dual certificate and its "distance" to degeneracy, while ours depend on the sharpness parameter c which is not trivial to characterize at first glance. In Theorem 4.6, we will provide an estimation for c via the so-called *Source Identity* that can be computed numerically.

3.3 Strong minima and robust recovery

A natural question is to whether robust recovery is still possible if the sharp minima property is replaced by the weaker strong minima property. We here show that the answer is affirmative, but at the price of a slower rate of convergence.

Theorem 3.12 (Strong minima for robust recovery)**.** *Suppose that J is a non-negative lsc function which satisfies (3.2), and is Lipschitz continuous around x_0 with modulus L. If x_0 is the unique and strong solution of (1.2) with constant $\kappa > 0$, then for all $\delta > 0$ sufficiently small, the following statements hold:*

(i) *Any solution x_δ to problem (1.4) satisfies*

$$\|x_\delta - x_0\| \le 2\left(\frac{1}{\kappa}L\|\Phi^\dagger\|\delta + \|\Phi^\dagger\|^2\delta^2\right)^{\frac{1}{2}}. \tag{3.24}$$

(ii) *For any $c_1 > 0$ and $\mu = c_1\delta$, any minimizer x_μ to (1.5) obeys*

$$\|x_\mu - x_0\| \le \sqrt{\frac{c_1}{(1 - c_1\kappa\|\Phi^\dagger\|^2\delta)\kappa}}\left(\frac{1}{c_1} + L\|\Phi^\dagger\|\right)\delta^{\frac{1}{2}}. \tag{3.25}$$

Proof. (i) Since x_0 is a strong solution of (1.2) with constant $\kappa > 0$, there exists $\nu > 0$ such that

$$J(x) - J(x_0) \ge \frac{\kappa}{2}\|x - x_0\|^2 \quad \text{for all} \quad x \in \Phi^{-1}y_0 \cap \mathbb{B}_\nu(x_0). \tag{3.26}$$

Let $\bar{x}_\delta$ be the projection of x_δ onto $\Phi^{-1}y_0$. Since x_0 is the unique solution of problem (1.2), we argue as in the proof of Theorem 3.10(i) to show that for all δ small enough, one has $\bar{x}_\delta, x_\delta \in \mathbb{B}_{\min\{\varepsilon,\nu\}}(x_0)$ with ε, ν in (3.19) and (3.26). We then have

$$\begin{aligned}
L\|\bar{x}_\delta - x_\delta\| &\ge J(\bar{x}_\delta) - J(x_\delta) && \text{by (3.19)},\\
&\ge J(\bar{x}_\delta) - J(x_0) && \text{by optimality of } x_0,\\
&\ge \frac{\kappa}{2}\|\bar{x}_\delta - x_0\|^2 && \text{by (3.26)},\\
&= \frac{\kappa}{2}\left(\|x_\delta - x_0\|^2 - \|\bar{x}_\delta - x_\delta\|^2\right). && \text{by (2.3)}.
\end{aligned}$$

This together with (2.3) again tells us that

$$
\begin{aligned}
\frac{\kappa}{2}\|x_\delta - x_0\|^2 &\le L\|\bar{x}_\delta - x_\delta\| + \frac{\kappa}{2}\|\bar{x}_\delta - x_\delta\|^2 \\
&= L\|\Phi^\dagger\Phi(x_\delta - x_0)\| + \frac{\kappa}{2}\|\Phi^\dagger\Phi(x_\delta - x_0)\|^2 \\
&\le L\|\Phi^\dagger\|\|\Phi x_\delta - y_0\| + \frac{\kappa}{2}\|\Phi^\dagger\|^2\|\Phi x_\delta - y_0\|^2 \\
&\le L\|\Phi^\dagger\|2\delta + \frac{\kappa}{2}\|\Phi^\dagger\|^2 4\delta^2,
\end{aligned}
$$

which is (3.24).

(ii) Let $\bar{x}_\mu$ be the projection of x_μ to $\Phi^{-1}y_0$. Arguing as above, we infer that for δ sufficiently small, $x_\mu, \bar{x}_\mu \in \mathbb{B}_{\min\{\varepsilon,\nu\}}(\bar{x})$, where ε, η are those in (3.19) and (3.26). Set for short $\alpha = \|\Phi(x_\mu - x_0)\|$. We then have

$$
\begin{aligned}
J(x_\mu) - J(x_0) &= J(\bar{x}_\mu) - J(x_0) + J(x_\mu) - J(\bar{x}_\mu) \\
&\ge \frac{\kappa}{2}\|\bar{x}_\mu - x_0\|^2 - L\|\bar{x}_\mu - x_\mu\| && \text{by (3.26) and (3.19),} \\
&= \frac{\kappa}{2}\left(\|x_\mu - x_0\|^2 - \|\bar{x}_\mu - x_\mu\|^2\right) - L\|\bar{x}_\mu - x_\mu\| && \text{by (2.3),} \\
&= \frac{\kappa}{2}\|x_\mu - x_0\|^2 - \|\Phi^\dagger\Phi(x_\mu - x_0)\|^2 - L\|\Phi^\dagger\Phi(x_\mu - x_0)\| && \text{by (2.3),} \\
&\ge \frac{\kappa}{2}\left(\|x_\mu - x_0\|^2 - \|\Phi^\dagger\|^2\alpha^2\right) - L\|\Phi^\dagger\|\alpha.
\end{aligned}
$$

Combining this with (3.22), we arrive at

$$
\frac{\kappa}{2}\left(\|x_\mu - x_0\|^2 - \|\Phi^\dagger\|^2\alpha^2\right) - L\|\Phi^\dagger\|\alpha \le \frac{2\alpha\delta - \alpha^2}{2\mu}.
$$

Making δ smaller if necessary so that $\mu = c_1\delta < \left(\kappa\|\Phi^\dagger\|^2\right)^{-1}$, we get

$$
\begin{aligned}
\frac{\kappa}{2}\|x_\mu - x_0\|^2 &\le \left(\frac{\delta}{\mu} + L\|\Phi^\dagger\|\right)\alpha - \left(\frac{1}{2\mu} - \frac{\kappa}{2}\|\Phi^\dagger\|^2\right)\alpha^2 \\
&\le \frac{1}{2}\left(\frac{1}{\mu} - \kappa\|\Phi^\dagger\|^2\right)^{-1}\left(\frac{\delta}{\mu} + L\|\Phi^\dagger\|\right)^2,
\end{aligned}
$$

which is (3.24) after some simple algebra.

□

It is worth noting that the upper bound in (3.25) is greater than or equal to the one in (3.24). Indeed, observe from the Young's inequality that

$$
\begin{aligned}
\sqrt{\frac{c_1}{(1 - c_1\kappa\|\Phi^\dagger\|^2\delta)\kappa}}\left(\frac{1}{c_1} + L\|\Phi^\dagger\|\right)\delta^{\frac{1}{2}} &\ge \sqrt{\frac{c_1}{(1 - c_1\kappa\|\Phi^\dagger\|^2\delta)\kappa}}2\sqrt{\frac{1}{c_1} - \kappa\|\Phi^\dagger\|^2\delta}\sqrt{L\|\Phi^\dagger\| + \kappa\|\Phi^\dagger\|^2\delta}\,\delta^{\frac{1}{2}} \\
&= 2\sqrt{\frac{1}{\kappa}L\|\Phi^\dagger\|\delta + \|\Phi^\dagger\|^2\delta^2},
\end{aligned}
$$

which is exactly the upper bound in (3.24). Equality holds with the choice $c_1 = \dfrac{1}{L\|\Phi^\dagger\| + 2\kappa\|\Phi^\dagger\|^2\delta}$.

When the regularizer J is convex, the Lipschitz property of J around x_0 is equivalent to the continuity of J at x_0; see, e.g., [3, Theorem 8.29]. Moreover, the assumption that x_0 is the unique minimizer of (1.2) holds trivially when x_0 is a strong minimizer. We then obtain the following corollary of Theorem 3.12.

Corollary 3.13. *Suppose that J is a convex function that is continuous at x_0. If x_0 is a strong solution of problem (1.2) then, for δ sufficiently small, the convergence rate $\mathcal{O}(\delta^{\frac{1}{2}})$ of Theorem 3.12 holds.*

Remark 3.14. *Observe that unlike sharp minima, even for the convex case, the property of strong minima is only local. This is the reason we still require δ to be small enough in Corollary 3.13. This is in stark contrast to Corollary 3.11 where sharpness is globalized under convexity. However, in the next result we provide an explicit bound for noise error δ at which estimates (3.24) and (3.25) occur for a large class of* convex piecewise linear-quadratic functions*. This class covers several important regularizers such as the ℓ_1 norm, the anisotropic total variation, the* elastic net regularizer *[47], the $\ell_1 + \ell_2$* regularizer *[22], and the* discrete Blake-Zisserman regularizer *[7].*

Recall from [35, Definition 10.20] that a function $J : \mathbb{R}^n \to \overline{\mathbb{R}}$ is called *piecewise linear-quadratic* if dom J is the union of finitely many polyhedral sets, relative to each of which $J(x)$ has the expression $\frac{1}{2}x^\top Ax + b^\top x + c$, where $A \in \mathbb{R}^{n\times n}$ is a symmetric matrix, $b \in \mathbb{R}^n$, and $c \in \mathbb{R}$. We say J to be convex piecewise linear-quadratic if it is convex and piecewise linear-quadratic. For this class of functions, according to [23, Theorem 2.7], the following global error bound property holds: for any $\gamma > 0$ there exists some $\kappa > 0$ such that

$$J(x) - J(x_0) \geq \frac{\kappa}{2}\|x - x_0\|^2 \quad \text{for} \quad x \in \mathbb{R}^n \quad \text{satisfying} \quad \Phi x = y_0 \text{ and } J(x) - J(x_0) \leq \gamma, \tag{3.27}$$

if additionally x_0 is a unique solution of problem (1.2).

Corollary 3.15. *Suppose that J is a continuous and convex piecewise linear-quadratic. If x_0 is a unique solution of (1.2), then*

(a) Statement (i) of Theorem 3.10 hold for any $\delta \leq \dfrac{\gamma}{2L\|\Phi^\dagger\|}$

(b) Statement (ii) of Theorem 3.10 hold for any $\delta < \min\left\{\dfrac{2c_1\gamma}{(1 + c_1L\|\Phi^\dagger\|)^2}, \dfrac{1}{c_1\kappa\|\Phi^\dagger\|^2}\right\}$

with $\gamma, \kappa > 0$ from (3.27) and some Lipschitz constant $L > 0$ of J.

Proof. It is similar to the proof of Corollary 3.11, there exists some $R > 0$ such that $x_0, x_\delta, \bar{x}_\delta, x_\mu, \bar{x}_\mu \in \mathbb{B}_R(0)$ and J is Lipschitz continuous with constant L on $\mathbb{B}_{2R}(0)$. Since x_0 is a unique solution of (1.2), it is also a strong solution due to (3.27). To justify (a), note that

$$\begin{aligned}
J(\bar{x}_\delta) - J(x_0) &= J(\bar{x}_\delta) - J(x_\delta) + J(x_\delta) - J(x_0) \\
&\leq L\|\bar{x}_\delta - x_\delta\| \\
&= L\|\Phi^\dagger\Phi(x_\delta - x_0)\| \\
&\leq L\|\Phi^\dagger\| \cdot \|\Phi(x_\delta - x_0)\| \\
&\leq L\|\Phi^\dagger\| \cdot (\|\Phi x_\delta - y\| + \|y - \Phi x_0)\|) \\
&\leq 2L\|\Phi^\dagger\|\delta.
\end{aligned}$$

It follows that $J(\bar{x}_\delta) - J(x_0) \le \gamma$ whenever $\delta \le \frac{\gamma}{2L\|\Phi^\dagger\|}$. Due to (3.27), we have

$$J(\bar{x}_\delta) - J(x_0) \ge \frac{\kappa}{2}\|x_\delta - x_0\|^2.$$

Repeating the inequalities at the end of the proof for (i) in Theorem 3.10, we have the same estimate (3.24).

To verify (b), set $\alpha := \|\Phi(x_\mu - x_0)\|$, we observe from (3.22) that

$$\begin{aligned}
J(\bar{x}_\mu) - J(x_0) &= J(\bar{x}_\mu) - J(x_\mu) + J(x_\mu) - J(x_0) \\
&\le L\|\bar{x}_\mu - x_\mu\| + \frac{1}{2\mu}(2\delta\alpha - \alpha^2) \\
&= L\|\Phi^\dagger\Phi(x_\mu - x_0)\| + \frac{1}{2\mu}(2\delta\alpha - \alpha^2) \\
&\le L\|\Phi^\dagger\|\alpha + \frac{1}{2\mu}(2\delta\alpha - \alpha^2) \\
&\le \left(\frac{1}{c_1} + L\|\Phi^\dagger\|\right)\alpha - \frac{1}{2\mu}\alpha^2 \\
&\le \frac{\mu}{2}\left(\frac{1}{c_1} + L\|\Phi^\dagger\|\right)^2.
\end{aligned}$$

As $\mu = c_1\delta$, the latter tells us that $J(\bar{x}_\mu) - J(x_0) \le \gamma$ whenever $\delta \le \frac{2c_1\gamma}{(1 + c_1L\|\Phi^\dagger\|)^2}$. By (3.27), we also have

$$J(\bar{x}_\mu) - J(\bar{x}) \ge \frac{\kappa}{2}\|\bar{x}_\mu - x_0\|^2.$$

Following the lines of the proof for (ii) in Theorem 3.10, we also obtain (3.25). Note that $\delta < \frac{1}{c_1\kappa\|\Phi^\dagger\|^2}$ is needed in that proof. □

Remark 3.16. *A natural and open question is whether the rate $\mathcal{O}(\delta^{\frac{1}{2}})$ in Theorem 3.12 is optimal under the strong solution property. Though we do not have a clear answer yet, we believe that this may depend on the regularizer J. For the case of Euclidean norm $J(x) = \|x\|$, we could prove that the rate could be improved to $\mathcal{O}(\delta)$. We omit the details here for the sake of brevity.*

4 Nondegeneracy, restricted injectivity and sharp minima

As shown in Proposition 3.8, sharp minima can be characterized via the descent and critical cones. However, it is not easy to verify them numerically, especially when the dimension is large[4]. In this section, we mainly derive quantitative characterizations for sharp minima to problem (1.2).

Throughout this section, we suppose that J takes the analysis-type form

$$J = J_0 \circ D^*, \tag{4.1}$$

where $D^* : \mathbb{R}^n \to \mathbb{R}^p$ is linear, and J_0 is a non-negative, convex and continuous function. The reason we take this form is twofold. First, this is in preparation for Section 5 to make the presentation easier there. Second, though we could derive the decomposability properties (see shortly)

[4] For some operators Φ drawn from some appropriate random ensembles, one can show that sample complexity bounds in [11, 39] are sufficient for sharpness to hold with high probability.

of J from those of J_0 using the framework in [43], our analysis in this section will involve a dual multiplier which is not the same as the one in that paper.

For some linear subspace V of $\mathbb{R}^n$, we will use the shorhand notation $w_V = \mathrm{P}_V w$ for $w \in \mathbb{R}^n$ and $D_V = D\mathrm{P}_V$ for a linear operator D.

4.1 Subdifferential decomposability

The following definition taken from [43, Definition 3] is instrumental in our study.

Definition 4.1 (Model Tangent Subspace). *Denote the* model vector *e of J_0 at $u_0 \stackrel{\text{def}}{=} D^* x_0$ as the projection of* 0 *onto the affine hull of $\partial J_0(u_0)$*

$$e \stackrel{\text{def}}{=} \mathrm{P}_{\mathrm{aff}\,(\partial J_0(u_0))}(0). \tag{4.2}$$

The model tangent subspace *T at u_0 associated to J_0 is defined by*

$$T \stackrel{\text{def}}{=} S^{\perp} \quad \textit{where} \quad S \stackrel{\text{def}}{=} \mathrm{aff}\,(\partial J_0(u_0)) - e. \tag{4.3}$$

Obviously, $e \in T$.

Lemma 4.2 (Decomposability, [43, Theorem 1]). *Let $u_0 \in \mathbb{R}^n \setminus \{0\}$ and v_0 be a vector in $\mathrm{ri}\,\partial J_0(u_0)$. Then*

$$\partial J_0(u_0) = \{v \in \mathbb{R}^n |\ v_T = e, \gamma_C(\mathrm{P}_S(v - v_0)) \leq 1\}, \tag{4.4}$$

where $C \stackrel{\text{def}}{=} \partial J_0(u_0) - v_0$ and γ_C is the gauge function of C defined in (2.7). Moreover $v \in \mathrm{ri}\,\partial J_0(u_0)$ if and only if $v_T = e$ and $\gamma_C(\mathrm{P}_S(v - v_0)) < 1$.

We then have the following necessary and sufficient optimality condition for problem (1.2).

Lemma 4.3. *x_0 is an optimal solution of problem (1.2) with J in (4.1) if and only if there exists $v \in \mathbb{R}^p$ such that*

$$v_T = e, \gamma_C(v_S - \mathrm{P}_S v_0) \leq 1, \ \textit{and} \ Dv \in \mathrm{Im}\,\Phi^*. \tag{4.5}$$

Proof. Combine (3.6) and Lemma 4.2. □

4.2 Quantitative characterization of sharp minima

We start by providing a quantitative condition for checking optimality.

Proposition 4.4 (Quantitative characterization of optimality). *x_0 is an optimal solution of problem (1.2) with J in (4.1) if and only if the following two conditions are satisfied*

$$ND_S(ND_S)^{\dagger} ND_T e = ND_T e \ \textit{and} \tag{4.6}$$

$$\rho(e) \stackrel{\text{def}}{=} \min_{u \in \mathrm{Ker}\, ND_S} \gamma_C(-(ND_S)^{\dagger} ND_T e - \mathrm{P}_S v_0 + u) \leq 1, \tag{4.7}$$

where N is a matrix satisfying $\mathrm{Ker}\, N = \mathrm{Im}\,\Phi^*$ and $\mathrm{Ker}\, N^* = \{0\}$. The minimum in (4.7) is well-defined and achieved at some $u \in \mathrm{Ker}\, ND_S \cap S$.

Proof. Suppose that x_0 is a solution of (1.2). It follows from Lemma 4.3 that there exists $v \in \mathbb{R}^p$ such that

$$0 = NDv = N(D_T e + D_S v_S) \Longrightarrow ND_S v_S = -ND_T e.$$

This is equivalent, via (2.2), to

$$ND_S(ND_S)^{\dagger} ND_T e = ND_T e \;\text{ and }\; v_S \in -(ND_S)^{\dagger} ND_T e + \operatorname{Ker} ND_S.$$

This tells us, using again Lemma 4.3, that

$$\rho(e) \leq \gamma_C(v_S - \mathrm{P}_S v_0) \leq 1,$$

which proves the necessary part.

Conversely, suppose that both (4.6) and (4.7) hold. Let $u^\star \in \operatorname{Ker} ND_S$ be a minimizer of (4.7). Since $\operatorname{dom} \gamma_C = S$ and γ_C is coercive on S by [43, Proposition 2], $u^\star$ exists and belongs to S. Define $\bar{u} \stackrel{\text{def}}{=} -(ND_S)^{\dagger} NDe + u^\star$. This vector verifies

$$\gamma_C(\bar{u} - \mathrm{P}_S v_0) = \rho(e) \leq 1. \tag{4.8}$$

Since $0 \in \operatorname{ri} C \subset C$, and in view of (4.8), we infer that $\bar{u} - \mathrm{P}_S v_0 \in C \subset S$. This implies $\bar{u} \in S$. Note further from (4.6) that

$$ND(e + \bar{u}) = ND(e + \mathrm{P}_S \bar{u}) = ND_T e - ND_S(ND_S)^{\dagger} ND_T e = 0,$$

which means $D(e + \bar{u}) \in \operatorname{Im} \Phi^*$. Altogether, the vector $e + \bar{u}$ verifies the properties (4.5) in Lemma 4.3 whence we deduce that x_0 is an optimal solution of problem (1.2). The proof is complete. □

Equation (4.6) means that the linear system

$$ND_S v = -ND_T e$$

is consistent. When $D = \mathbb{I}$, Lemma 4.3 and Proposition 4.4 tell us that conditions (4.6)-(4.7) are equivalent to the so-called Source Condition well-known in inverse problems (see [37, 20, 42, 45] and references therein). We call $\rho(e)$ the *Source Coefficient*. Under condition (4.6), $\rho(e)$ is indeed the optimal value to the following problem

$$\min_{v \in \mathbb{R}^p} \gamma_C(v - \mathrm{P}_S v_0) \quad \text{subject to} \quad ND_S v = -ND_T e,$$

which is equivalent to the following convex optimization problem

$$\min_{v \in \mathbb{R}^p} \gamma_C(v - \mathrm{P}_S v_0) \quad \text{subject to} \quad NDv = -ND_T e \;\text{ and }\; v \in S. \tag{4.9}$$

Remark 4.5 (Computing the Source Coefficient $\rho(e)$ for decomposable norms)**.** *One important class of regularizers is $J(x) = \|D^* x\|_{\mathcal{A}}$, where $J_0(\cdot) = \|\cdot\|_{\mathcal{A}}$ is a* decomposable *norm in $\mathbb{R}^p$. The class of decomposable norms is introduced in [29, 8] and was generalized in [43] (coined strong gauges there). This class includes the ℓ_1 norm, the ℓ_1/ℓ_2 norm, the nuclear norm but not the ℓ_∞ norm. According to [8], a norm $\|\cdot\|_{\mathcal{A}}$ is called decomposable at $u_0 \stackrel{\text{def}}{=} D^* x_0$ if there is a subspace $V \subset \mathbb{R}^p$ and a vector $e_0 \in V$ such that*

$$\partial \|u_0\|_{\mathcal{A}} = \{v \in \mathbb{R}^p |\; v_V = e_0 \;\textit{and}\; \|v_{V^\perp}\|_{\mathcal{A}}^* \leq 1\}, \tag{4.10}$$

where $\|\cdot\|_{\mathcal{A}}^*$ is the dual norm to $\|\cdot\|_{\mathcal{A}}$. From [43, Proposition 7], it follows that (4.10) complies with Lemma 4.2 by taking $V = T = S^\perp = (\mathrm{aff}\,(\partial J_0(u_0)) - e)^\perp$, $v_0 = e \in \mathrm{ri}\,\partial\|u_0\|_{\mathcal{A}}$, in which case $C = \partial\|u_0\|_{\mathcal{A}} - v_0 = \{v \in S |\ \|v\|_{\mathcal{A}}^* \le 1\}$ and thus for any $v \in S$

$$\gamma_C(v) = \|v\|_{\mathcal{A}}^*.$$

Hence, problem (4.9) simplifies to

$$\min_{v\in\mathbb{R}^p} \|v\|_{\mathcal{A}}^* \quad \text{subject to} \quad NDv = -ND_T e \ \text{ and } \ v \in S. \tag{4.11}$$

The matrix N can be chosen from the singular value decomposition of $\Phi = U\Sigma V^*$ as $N = V_G^*$, where V_G is the submatrix of V whose columns are indexed by $G = \{r+1, r+2, \ldots, n\}$ with $r = \mathrm{rank}\,\Phi$. This idea is slightly inspired from [28, 48] for the case where $\|\cdot\|_{\mathcal{A}}$ is the ℓ_1 norm. We will return in Section 5 to discussing further the use of the Source Coefficient to classify sharp and strong/unique solutions for the group-sparsity regularization.

We are now in position to state the main result of this section, which provides equivalent characterizations of sharp minima.

Theorem 4.6 (Characterizations of sharp solutions)**.** *Consider J in (4.1). The following statements are equivalent:*

(i) *x_0 is a sharp solution of problem (1.2).*

(ii) *The following* Restricted Injectivity *holds at x_0*

$$\mathrm{Ker}\,\Phi \cap \mathrm{Ker}\,D_S^* = \{0\} \tag{4.12}$$

and the Source Coefficient *$\rho(e) < 1$.*

(iii) *The Restricted Injectivity (4.12) holds and the* Nondegenerate Source Condition *is satisfied at x_0 in the sense that*

$$\exists v \in \mathbb{R}^p \quad \text{such that} \quad Dv \in \mathrm{Im}\,\Phi^* \ \text{ and } \ v \in \mathrm{ri}\,\partial J_0(x_0). \tag{4.13}$$

Moreover, the sharpness constant c at x_0 can be as large as

$$c = (1-\rho(e))sc_1 > 0 \quad \text{with} \quad c_1 \overset{\text{def}}{=} \min_{w\in\mathrm{Ker}\,\Phi\cap\mathbb{S}^{n-1}} \|D_S^* w\| > 0 \tag{4.14}$$

and $s > 0$ satisfying $\mathbb{B}_s(0) \cap S \subset C$.

Proof. [(i)⇒(ii)]. Suppose that x_0 is a sharp solution of (1.2). From Lemma 2.4(i), this is equivalent to the existence of some $c > 0$ such that $d\Psi(x_0)(w) \ge c\|w\|$ for all $w \in \mathbb{R}^n$. From (3.6) and Lemma 4.2, we have

$$\begin{aligned}
d\Psi(x_0)(w) &= dJ(x_0)(w) + \iota_{\mathrm{Ker}\,\Phi}(w)\\
&= dJ_0(D^*x_0)(D^*w) + \iota_{\mathrm{Ker}\,\Phi}(w)\\
&= \sigma_{\partial J_0(D^*x_0)}(D^*w) + \iota_{\mathrm{Ker}\,\Phi}(w)\\
&= \sup\{\langle v,\, D^*w\rangle \mid v_T = e, v_S \in \mathrm{P}_S v_0 + C\} + \iota_{\mathrm{Ker}\,\Phi}(w)\\
&= \langle e,\, D_T^*w\rangle + \langle \mathrm{P}_S v_0,\, D_S^*w\rangle + \sup_{z\in C}\langle z,\, D_S^*w\rangle + \iota_{\mathrm{Ker}\,\Phi}(w)\\
&= \langle e,\, D_T^*w\rangle + \langle \mathrm{P}_S v_0,\, D_S^*w\rangle + \sigma_C(D_S^*w) + \iota_{\mathrm{Ker}\,\Phi}(w).
\end{aligned} \tag{4.15}$$

For any $w = N^*u \in \operatorname{Ker}\Phi$, we obtain from (4.15), the inequality $d\Psi(x_0)(w) \geq c\|w\|$, and (4.6) that

$$\begin{aligned}
\sigma_C(D_S^* w) - c\|w\| &\geq -\langle e,\, D_T^* N^* u\rangle - \langle \mathrm{P}_S v_0,\, D_S^* N^* u\rangle \\
&= -\langle ND_T e,\, u\rangle - \langle \mathrm{P}_S v_0,\, D_S^* N^* u\rangle \\
&= -\left\langle ND_S (ND_S)^\dagger ND_T e,\, u\right\rangle - \langle \mathrm{P}_S v_0,\, D_S^* N^* u\rangle \\
&= -\left\langle (ND_S)^\dagger ND_T e + \mathrm{P}_S v_0,\, D_S^* w\right\rangle .
\end{aligned} \tag{4.16}$$

If $w \in \operatorname{Ker}\Phi \cap \operatorname{Ker} D_S^*$, we deduce from (4.16) that $-c\|w\| \geq 0$, which means $w = 0$. Thus the Restricted Injectivity (4.12) is satisfied.

Moreover, as J_0 is a continuous convex function, $\partial J_0(D^* x_0)$ is compact, and so is C, whence it follows that

$$\sigma_C(D_S^* w) \leq r\|D_S^* w\| \leq r\|D_S^*\|\|w\| \quad \text{with} \quad r \stackrel{\text{def}}{=} \max\left\{\|z\| \,|\, z \in C\right\}.$$

Combining the latter with (4.16) tells us that for any $v \in \operatorname{Im} D_S^* N^*$,

$$-\left\langle (ND_S)^\dagger ND_T e + \mathrm{P}_S v_0,\, v\right\rangle \leq \left(1 - \frac{c}{r\|D_S^*\|}\right)\sigma_C(v). \tag{4.17}$$

Let C° be the polar set of C as defined in (2.9), and set $K \stackrel{\text{def}}{=} C^\circ \cap S$. K is a non-empty closed convex set. Since $0 \in \operatorname{ri} C$, there exists $s > 0$ such that and $\mathbb{B}_s(0) \cap S \subset C$. For any $v \in K \setminus \{0\}$, we have $s\frac{v}{\|v\|} \in \mathbb{B}_s(0) \cap S$ and thus, using (2.9), $\left\langle v,\, s\frac{v}{\|v\|}\right\rangle \leq \sigma_{\mathbb{B}_s(0)\cap S}(v) \leq \sigma_C(v) \leq 1$ since $v \in C^\circ$. It follows that $\|v\| \leq \frac{1}{s}$. Hence, K is a compact set.

Let us bound from below the right hand side of (4.17). First, we have by Fenchel-Moreau theorem and the minimax theorem [33, Corollary 37.3.2] (since K is compact), that

$$\begin{aligned}
\max_{v \in K \cap \operatorname{Im} D_S^* N^*} \left\langle -(ND_S)^\dagger ND_T e - \mathrm{P}_S v_0,\, v\right\rangle &= \max_{v \in K}\left(\left\langle -(ND_S)^\dagger ND_T e - \mathrm{P}_S v_0,\, v\right\rangle - \iota_{\operatorname{Im} D_S^* N^*}(v)\right) \\
&= \max_{v \in K}\inf_{z \in \operatorname{Ker} ND_S}\left\langle -(ND_S)^\dagger ND_T e - \mathrm{P}_S v_0 + z,\, v\right\rangle \\
&= \min_{z \in \operatorname{Ker} ND_S}\max_{v \in K}\left\langle -(ND_S)^\dagger ND_T e - \mathrm{P}_S v_0 + z,\, v\right\rangle \\
&= \min_{z \in \operatorname{Ker} ND_S}\sigma_K(-(ND_S)^\dagger ND_T e - \mathrm{P}_S v_0 + z).
\end{aligned} \tag{4.18}$$

On the other hand, for any $v \in K \cap \operatorname{Im} D_S^* N^* \subset C^\circ$, we have $\sigma_C(v) \leq 1$ (see (2.9)). Let $v^\star \in K \cap \operatorname{Im} D_S^* N^*$ be a maximizer of the left hand side of (4.18). We then have from (4.17) and (4.18) that

$$\begin{aligned}
0 \leq \min_{z \in \operatorname{Ker} ND_S}\sigma_K(-(ND_S)^\dagger ND_T e - \mathrm{P}_S v_0 + z) &= -\left\langle (ND_S)^\dagger ND_T e + \mathrm{P}_S v_0,\, v^\star\right\rangle \\
&\leq \left(1 - \frac{c}{r\|D_S^*\|}\right)\sigma_C(v^\star) \leq 1 - \frac{c}{r\|D_S^*\|} < 1.
\end{aligned} \tag{4.19}$$

It remains now to compute σ_K. We have from Fenchel-Moreau theorem, [33, Corollary 16.4.1], standard conjugacy calculus and (2.10), that

$$\sigma_K = (\iota_{C^\circ} + \iota_S)^* = \operatorname{cl}\left(\inf_{\eta \in \mathbb{R}^p}\sigma_{C^\circ}(\cdot + \eta) + \sigma_S(\eta)\right) = \operatorname{cl}\left(\inf_{\eta \in T}\sigma_{C^\circ}(\cdot + \eta)\right) = \operatorname{cl}\left(\inf_{\eta \in T}\gamma_C(\cdot + \eta)\right).$$

Since $\operatorname{dom}\gamma_C = S$ by [43, Proposition 2], we have for any $z \in \mathbb{R}^p$ that

$$\inf_{\eta \in T} \gamma_C(z + \eta) = \inf_{\eta \in T} \gamma_C(\mathrm{P}_S z + \eta) = \gamma_C(\mathrm{P}_S z).$$

Thus, the closure operation can be omitted above to get for any $z \in S$

$$\sigma_K(z) = \gamma_C(z).$$

Inserting this into (4.19), we get (ii), after observing that $\operatorname{Im}(ND_S)^\dagger = \operatorname{Im} D_S^* N^* \subset S$ and that the minimum in $\rho(e)$ is achieved on S (see Proposition 4.4).

[(ii)⇒(iii)]. Let $u^\star \in S$ be a solution of the minimization problem in (4.7), denote $\bar{u} = -(ND_S)^\dagger NDe + u^* \in S$ and $\bar{v} = e + \bar{u}$. We have $\bar{v}_T = e$ and, since (ii) is satisfied, $\gamma_C(\bar{v}_S - \mathrm{P}_S v_0) < 1$. This is equivalent to $\bar{v} \in \operatorname{ri} \partial J(x_0)$ thanks to the last claim of Lemma 4.2. It remains to show that $D\bar{v} \in \operatorname{Im} \Phi^*$. As the Restricted Injectivity condition (4.12) holds,

$$ND_S(ND_S)^\dagger ND_T e = ND_S(ND_S)^*(ND_S(ND_S)^*)^{-1} ND_T e = ND_T e,$$

which verifies (4.6). (4.7) is also verified as shown above and we can then argue as in the proof of the sufficient part in Proposition 4.4 to deduce that $D\bar{v} \in \operatorname{Im} \Phi^*$.

[(iii)⇒(i)]. Suppose that the Restricted Injectivity (4.12) holds at x_0 and there exists $Dv \in \operatorname{Im} \Phi^*$ and $v \in \operatorname{ri} \partial J(x_0)$. Hence, $v_T = e$ and $\gamma_C(\mathrm{P}_S(v - v_0)) < 1$ thanks to Lemma 4.2. Thus, for any $w \in \operatorname{Ker} \Phi$, we get from (4.15) that

$$\begin{aligned} d\Psi(x_0)(w) &= \langle e, D^* w\rangle + \langle \mathrm{P}_S v_0, D_S^* w\rangle + \sigma_C(D_S^* w) \\ &= \langle v - \mathrm{P}_S v, D^* w\rangle + \langle \mathrm{P}_S v_0, D_S^* w\rangle + \sigma_C(D_S^* w) \\ &= \langle -\mathrm{P}_S v, D^* w\rangle + \langle \mathrm{P}_S v_0, D_S^* w\rangle + \sigma_C(D_S^* w) \\ &= \langle -\mathrm{P}_S(v - v_0), D_S^* w\rangle + \sigma_C(\mathrm{P}_S w) \\ &\geq -\sigma_C(D_S^* w)\gamma_C(\mathrm{P}_S(v - v_0)) + \sigma_C(D_S^* w) \\ &= \Big(1 - \gamma_C(\mathrm{P}_S(v - v_0))\Big)\sigma_C(D_S^* w), \end{aligned} \tag{4.20}$$

where in the last two inequalities, we used the duality inequality on $\operatorname{dom}\sigma_C \times \operatorname{dom}\gamma_C = \mathbb{R}^p \times \operatorname{dom} S$. Recall that $0 \in \operatorname{ri} C$ and $\mathbb{B}_s(0) \cap S \subset C$ for some $s > 0$. As $D_S^* w \in S$, we have

$$\sigma_C(D_S^* w) \geq \left\langle D_S^* w, s\frac{D_S^* w}{\|D_S^* w\|} \right\rangle = s\|D_S^* w\| \geq s c_1 \|w\|,$$

where $c_1 \stackrel{\text{def}}{=} \min\left\{\|D_S^* w\| \mid w \in \mathbb{S}^{n-1} \cap \operatorname{Ker}\Phi\right\}$, and $c_1 > 0$ by virtue of the Restricted Injectivity (4.12). We derive from (4.20) that

$$d\Psi(x_0)(w) \geq \Big(1 - \gamma_C(\mathrm{P}_S(v - v_0))\Big) s c_1 \|w\| \quad \text{for all} \quad w \in \operatorname{Ker}\Phi,$$

which verifies (i) by Lemma 2.4. Recall that $Dv \in \operatorname{Im}\Phi^*$ and $v_T = e$, and thus $NDv_S = ND_T e$, which in turn shows that the vector v_S obeys the constraint in (4.9), and thus $\rho(e) \leq \gamma_C(\mathrm{P}_S(v - v_0)) < 1$. The sharpness constant of Ψ can then be as large as devised in (4.14). The proof is complete. □

Remark 4.7. *The sharpness constant c in (4.14) is based on the Source Coefficient $\rho(e)$ and the constant c_1 which captures the* Restricted Injectivity *property (4.12). The computation of $\rho(e)$ was discussed in (4.9) and (4.11). The constant c_1 is the minimum singular value of D_S^* restricted to* Ker Φ, *which is a reminiscent* *of the so-called minimum conic singular value of a matrix. Observe that $c_1 > 0$ thanks to the validity of* (4.12)*. Moreover, we can compute c_1^2 easily as the optimal value of the quadratic program*

$$\min_{w\in\mathbb{R}^n} \quad w^\top D_S D_S^* w \quad \textit{subject to} \quad w^\top w \le 1 \quad \textit{and} \quad \Phi w = 0,$$

which can be solved by available packages such as `cvxopt`*.*

Corollary 4.8 (Sufficient condition for sharp solution)**.** *If the Restricted Injectivity holds at x_0 and the following condition*

$$\tau(e) \stackrel{\text{def}}{=} \gamma_C(-(ND_S)^\dagger NDe - \mathrm{P}_S v_0) < 1 \tag{4.21}$$

is satisfied, then x_0 is a sharp solution of problem (1.2)*.*

Proof. It is easy to see that $\tau(e) \ge \rho(e)$. The result follows from Theorem 4.6. ☐

Remark 4.9. *The Restricted Injectivity* (4.12) *was proposed in [15, 42, 43, 45]. It also has tracks in some special cases, e.g., for the ℓ_1 norm problems [17, 18, 20, 28, 44, 49, 48], ℓ_1/ℓ_2 norm problems [19, 24], and nuclear norm problems [11, 9, 8]. The form of the Nondegenerate Source Condition in Theorem 4.6 appears also in [42, 43, 45]. It generalizes the condition in [20] for the ℓ_1 problem, which actually occurred earlier in [10]. The combination of Restricted Injectivity and Nondegenerate Source Condition in the above result are proved in [42, Theorem 5.3] as sufficient conditions for solution uniqueness to problem* (1.2)*, but they are not necessary in general. By revisiting Example 3.9, we see that x_0 is a unique solution of problem* (3.16) *but the Nondegenerate Source Condition is not satisfied. Indeed, the only vector $v \in \mathbb{R}^2$ satisfying $\Phi^* v = (v_1+v_2, v_1, -v_2)^\top \in \partial J(x_0) = \{(0,1)\} \times [-1,1]$ is $v = (1,-1)$, but $\Phi^* v = (0,1,1) \notin \mathrm{ri}\, \partial J(x_0)$.*

Theorem 4.6 strengthens [43, Corollary 1] by showing Restricted Injectivity and Nondegenerate Source Condition are necessary and sufficient for sharp minima. In the case of ℓ_1 problem, out result recovers part of [48, Theorem 2.1] that gives a characterization for solution uniqueness, which is equivalent to sharp minima in this framework. Some other characterizations are also studied recently in [18, 27] by exploiting polyhedral structures. It is worth emphasizing that our result does not need polyhedrality. Our proof mainly relies on the well-known first-order condition in Lemma 2.4 for sharp minima and the subdifferential decomposiability (4.4)*. Theorem 4.6 also covers many results about solution uniqueness in [11, 9, 8, 10, 17, 19, 18, 20, 24, 28, 36, 44, 40, 41, 49, 48].*

Remark 4.10. *The idea of using first-order analysis to study solution uniqueness to regularized optimization problems is not new as discussed above. For instance, the* Null Space Property *has been shown to ensure solution uniqueness for ℓ_1 regularization [13, 28]. A generalization of this condition beyond the ℓ_1 norm, coined* Strong Null Space Property*, was proposed in [15, 42, 43]. This property reads*

$$-\langle e, D_T^* w\rangle - \langle \mathrm{P}_S v_0, D_S^* w\rangle < \sigma_C(D_S^* w) \quad \textit{for all} \quad w \in \mathrm{Ker}\,\Phi \setminus \{0\}. \tag{4.22}$$

It is immediate to see from (4.16) *that sharpness at x_0 entails* (4.22)*.*

The case of analysis ℓ_1 When $\|\cdot\|_{\mathcal{A}} = \|\cdot\|_1$, the model vector e in (4.2) is indeed $(\mathrm{sign}\,(D_I^* x_0), 0_K)$, where $I \stackrel{\text{def}}{=} \mathrm{supp}\,(D^* x_0) \stackrel{\text{def}}{=} \{i \in \{1,2,\ldots,p\} |\ (D^* x_0)_i \neq 0\}$, $K \stackrel{\text{def}}{=} \{1,\ldots,p\} \setminus I$, and D_I is the submatrix of D with column indices I. In his case, inequality (4.21) takes the form

$$\tau(\mathrm{sign}\,(D_I^* x_0)) = \|(ND_K)^\dagger ND_I \mathrm{sign}\,(D_I^* x_0)\|_\infty < 1,$$

which is called the *Analysis Exact Recovery Condition* in [28]. Another criterion used in [44] to check solution uniqueness for ℓ_1 problem is the so-called *Analysis Identifiability Criterion* at $\mathrm{sign}\,(D_I^* x_0)$ denoted by

$$\mathbf{IC}(\mathrm{sign}\,(D_I^* x_0)) \stackrel{\text{def}}{=} \min_{v\in \mathrm{Ker}\, D_J} \|D_K^\dagger (\Phi^* (\Phi_U^\dagger)^* U^* - \mathbb{I}) D_I \mathrm{sign}\,(D_I^* x_0) - v\|_\infty < 1, \tag{4.23}$$

where U is basis of $\operatorname{Ker} D_J^*$ and $\Phi_U \stackrel{\text{def}}{=} \Phi U$. This condition reduces to the synthesis one introduced in [17] in the case of $D^* = \mathbb{I}$ while $\tau(\operatorname{sign}(D_I^* x_0))$ does not. As discussed in [28, 44], $\tau(\operatorname{sign}(D_I^* x_0))$ and $\mathbf{IC}(\operatorname{sign}(D_I^* x_0))$ are different and no one implies the other even for the case $D = \mathbb{I}$.

For our general framework with J as in (4.1), the Analysis Identifiability Criterion is satisfied at x_0 if

$$\mathbf{IC}(e) \stackrel{\text{def}}{=} \min_{u \in \operatorname{Ker} D_S} \gamma_C(D_S^\dagger(\Phi^*(\Phi_U^\dagger)^* U^* - \mathbb{I}) D_T e - \mathrm{P}_S v_0 + u) < 1, \tag{4.24}$$

where U is a matrix whose columns form a basis of $\operatorname{Ker} D_S^*$ and $\Phi_U \stackrel{\text{def}}{=} \Phi U$.

Proposition 4.11. *If the Restricted Injectivity (4.12) holds at x_0 then $\rho(e) \leq \mathbf{IC}(e)$. Consequently, if the Analysis Identifiability Criterion holds at x_0, then x_0 is a sharp solution of (1.2).*

Proof. (4.12) means that $\operatorname{Ker} \Phi_U = \{0\}$ and $\Phi_U^\dagger = (\Phi_U^* \Phi_U)^{-1} \Phi_U^*$. Let $\bar{u} \in \operatorname{Ker} D_S$ be a minimizer of problem (4.24), and define

$$\bar{v} \stackrel{\text{def}}{=} D_S^\dagger(\Phi^*(\Phi_U^\dagger)^* U^* - \mathbb{I}) D_T e + \bar{u} \quad \text{with} \quad \gamma_C(\bar{v} - \mathrm{P}_S v_0) = \mathbf{IC}(e).$$

Note that

$$U^*(\Phi^*(\Phi_U^\dagger)^* U^* - \mathbb{I}) D_T e = (\Phi_U^* \Phi_U (\Phi_U^* \Phi_U)^{-1} U^* - U^*) D_T e = 0.$$

It follows that $(\Phi^*(\Phi_U^\dagger)^* U^* - \mathbb{I}) D_T e \in \operatorname{Im} D_S$, which implies that $D_S \bar{v} = (\Phi^*(\Phi_U^\dagger)^* U^* - \mathbb{I}) D_T e$ by (2.2). Hence we have

$$N D_S \bar{v} = (N \Phi^* (\Phi_U^\dagger)^* U^* - N) D_T e = -N D_T e, \tag{4.25}$$

which implies that $\bar{v} \in -(N D_S)^\dagger N D_T e + \operatorname{Ker} N D_S$ and thus $\rho(e) \leq \gamma_C(\bar{v} - \mathrm{P}_S v_0) = \mathbf{IC}(e)$. When the Analysis Identifiability Criterion is satisfied at x_0, we have $\rho(e) \leq \mathbf{IC}(e) < 1$ and thus x_0 is a sharp solution of (1.2) due to Theorem 4.6. The proof is complete. □

In plain words, Proposition 4.11 tells us that $\rho(e) < 1$ is weaker than the Analysis Identifiability Criterion (4.24). In turn Theorem 4.6 is stronger than [44, Theorem 2] for the analysis ℓ_1 problem. It also covers the [42, Proposition 5.7]. For the analysis ℓ_1 problem, [48] provides an example where $\rho(\operatorname{sign}(D_I^* x_0))$ is strictly smaller than both $\tau(\operatorname{sign}(D_I^* x_0))$ and $\mathbf{IC}(\operatorname{sign}(D_I^* x_0))$.

4.3 Robust recovery with analysis decomposable priors

The following result proves robust recovery with linear rate under Restricted Injectivity and Non-degenerate Source Condition.

Corollary 4.12 (Robust recovery of decomposable norm minimization)**.** *Suppose that J_0 is a nonnegative continuous convex function satisfying*

$$(D^*)^{-1}(\operatorname{Ker}(J_0)_\infty) \cap \operatorname{Ker} \Phi = \{0\}. \tag{4.26}$$

If (4.12)-(4.13) hold at x_0, we have linear convergence rate for robust recovery as in (3.17) and (3.18).

Proof. From [2, Proposition 2.6.3] $J_\infty(w) = (J_0)_\infty(D^* w)$ for any $w \in \mathbb{R}^n$. Hence, condition (4.26) is exactly (3.2). We obtain the claim by combining Theorem 4.6 and Corollary 3.11. □

Corollary 4.12 covers parts of [15, Theorem 2] (see also [42, 45]), [20, Theorem 4.2], and [48, Theorem 2] for the case of (analysis or synthesis) ℓ_1 minimization problems. When $D^* = \mathbb{I}$ and $J_0(x) = \|x\|_{\mathcal{A}}$, it also covers the results in [11] about robust recovery for the constrained problem (1.4).

4.4 Connections between unique/sharp/strong solutions in the noiseless case

When there is noise in observation (1.3), problem (1.5) is usually used to recover the original signal x_0. Solution uniqueness to (1.5) is especially important for exact recovery [17, 44]. We show next that Restricted Injectivity and Nondegenerate Source Condition are sufficient for strong minima to problem (1.5), and they become necessary for the ℓ_1 problem. This result is true for a larger class of regularized problems taking the form

$$\min_{x\in\mathbb{R}^n} \quad \Theta(x) \stackrel{\text{def}}{=} f(\Phi x) + \mu J(x), \tag{4.27}$$

where μ is a positive parameter and the loss function $f : \mathbb{R}^m \to \overline{\mathbb{R}}$ is an extended real-valued convex function satisfying the following two conditions:

(A) f is twice continuously differentiable in $\operatorname{int}(\operatorname{dom} f)$.

(B) $\nabla f^2(x)$ is positive definite for all $x \in \operatorname{int}(\operatorname{dom} f)$, i.e., f is strictly convex in the interior of its domain.

In (1.5), the function f is $\frac{1}{2}\|\cdot - y\|^2$ which certainly satisfies the above two conditions. Moreover, the standing assumptions (A) and (B) for f cover the important case of the Kullback-Leiber divergence:

$$f_{\text{KL}}(z) = \begin{cases} \sum_{i=1}^{m} y_i \log \dfrac{y_i}{z_i} + z_i - y_i & \text{if} \quad z \in \mathbb{R}^m_{++} \\ +\infty & \text{if} \quad z \in \mathbb{R}^m_{+} \setminus \mathbb{R}^m_{++}, \end{cases} \tag{4.28}$$

where $y \in \mathbb{R}^m_+$ and $0 \log 0 = 0$. This offers a natural way to measure of similarity of two nonnegative vectors (e.g., two discrete distributions) and is broadly used in statistical/machine learning and signal processing.

In the following result, we provide the connections between unique/sharp/strong solutions for the two problems (1.2) and (4.27). Part (i) of this result could be obtained from [27, Proposition 3.2], but we still give a short proof as our assumptions on f are slightly different, e.g., f may not have full domain.

Proposition 4.13 (Unique/sharp/strong solutions for problems (1.2) and (4.27))**.** *Suppose that $\bar{x}$ is an optimal solution of problem (4.27) and $\Phi\bar{x} \in \operatorname{int}(\operatorname{dom} f)$. Then $\bar{x}$ is a solution of problem (1.2) with $x_0 = \bar{x}$. Moreover, the following statements hold:*

(i) *$\bar{x}$ is the unique solution of (4.27) if and only if it is the unique solution of (1.2).*

(ii) *If $\bar{x}$ is a sharp solution of (1.2) then it is a strong solution of problem (4.27).*

(iii) *If $\bar{x}$ is a strong solution of (4.27) then it is also a strong solution of (1.2).*

Proof. For any $x \in \Phi^{-1}(\Phi\bar{x})$, we have

$$f(\Phi\bar{x}) + \mu J(x) = f(\Phi x) + \mu J(x) \geq f(\Phi\bar{x}) + \mu J(\bar{x}). \tag{4.29}$$

It follows that $\bar{x}$ is a solution of problem (1.2) with $y_0 = \Phi\bar{x}$.

(i) If $\bar{x}$ is the unique solution of problem (4.27), we have the strict inequality in (4.29) provided that $x \neq \bar{x}$, which shows that $\bar{x}$ is also the unique solution of problem (1.2). Assume now that $\bar{x}$ is the unique solution of (1.2). By contradiction, suppose that $\bar{x}$ is not the unique solution of (4.27). Since Argmin (Θ) is convex and $\Phi^{-1}(\text{int}\,(\text{dom}\, f))$ is an open set containing $\bar{x}$, there exists $r > 0$ such that $\mathbb{B}_r(\bar{x}) \subset \Phi^{-1}(\text{int}\,(\text{dom}\, f))$ and $\mathbb{B}_r(\bar{x}) \cap \text{Argmin}\,(\Theta) \neq \emptyset$, while the later is not the singleton $\{\bar{x}\}$. Choose $\widehat{x} \in \mathbb{B}_r(\bar{x})$ with $\widehat{x} \neq \bar{x}$. We have $-\Phi^*\nabla f(\Phi\widehat{x}) \in \partial J(\widehat{x})$ and $-\Phi^*\nabla f(\Phi\bar{x}) \in \partial J(\bar{x})$. The monotonicity of the subdifferential tells us that

$$- \langle \nabla f(\Phi\widehat{x}) - \nabla f(\Phi\bar{x}), \Phi\widehat{x} - \Phi\bar{x} \rangle = - \langle \Phi^*\nabla f(\Phi\widehat{x}) - \Phi^*\nabla f(\Phi\bar{x}), \widehat{x} - \bar{x} \rangle \geq 0.$$

Convexity of f entails the opposite inequality which shows that

$$\langle \nabla f(\Phi\widehat{x}) - \nabla f(\Phi\bar{x}), \Phi\widehat{x} - \Phi\bar{x} \rangle = 0.$$

Since f is twice continuously differentiable in int $(\text{dom}\, f)$, we obtain from the mean-value theorem that

$$\int_0^1 \left\langle \nabla^2 f(\Phi(\bar{x} + t(\widehat{x} - \bar{x})))\Phi(\widehat{x} - \bar{x}), \Phi(\widehat{x} - \bar{x}) \right\rangle dt = 0. \tag{4.30}$$

Since the Hessian is positive definite and continuous on int $(\text{dom}\, f)$, there exists some $\tau > 0$ such that

$$\left\langle \nabla^2 f(\Phi(\bar{x} + t(\widehat{x} - \bar{x})))\Phi(\widehat{x} - \bar{x}), \Phi(\widehat{x} - \bar{x}) \right\rangle \geq \tau \|\Phi(\widehat{x} - \bar{x})\|^2$$

for all $t \in [0, 1]$. Combining this with (4.30) implies that $\Phi\widehat{x} = \Phi\bar{x}$. Using this together with the fact that both $\widehat{x}$ and $\bar{x}$ are minimizers to (4.27), entails that $J(\widehat{x}) = J(\bar{x})$. This contradicts uniqueness of $\bar{x}$ for (1.2).

(ii) Assume that $\bar{x}$ is a sharp solution of (1.2). By Proposition 3.8, we have

$$\text{Ker}\,\Phi \cap \mathcal{C}_J(x_0) = \{0\} \tag{4.31}$$

where we recall $\mathcal{C}_J(x_0)$ from (3.10). From the sum rule (2.17) and convexity of J, we have

$$d^2\Theta(\bar{x}|0)(w) = \left\langle \nabla^2 f(\Phi\bar{x})\Phi w, \Phi w \right\rangle + \mu d^2 J(\bar{x}| - \mu^{-1}\Phi^*\nabla f(\Phi\bar{x})(w)) \geq \left\langle \nabla^2 f(\Phi\bar{x})\Phi w, \Phi w \right\rangle \tag{4.32}$$

and from (2.15), dom $d^2\Theta(\bar{x}|0) \subset \{w \in \mathbb{R}^n |\ d\Theta(\bar{x})(w) = 0\}$. We also have

$$d\Theta(\bar{x})(w) = \langle \nabla f(\Phi\bar{x}), \Phi w \rangle + \mu dJ(x_0)(w).$$

It then follows that

$$\text{dom}\, d^2\Theta(\bar{x}|0) \subset \{w \in \mathbb{R}^n |\ \langle \nabla f(\Phi\bar{x}), \Phi w \rangle + \mu dJ(x_0)(w) = 0\}. \tag{4.33}$$

To verify that $\bar{x}$ is a strong solution of problem (4.27) by using Lemma 2.4, we claim that

$$\left\langle \nabla^2 f(\Phi\bar{x})\Phi w, \Phi w \right\rangle > 0 \quad \text{for all} \quad w \in \text{dom}\, d^2\Theta(\bar{x}|0) \setminus \{0\}. \tag{4.34}$$

Suppose that $w \in \text{dom}\, d^2\Theta(\bar{x}|0)$ satisfying $\left\langle \nabla^2 f(\Phi\bar{x})\Phi w, \Phi w \right\rangle \leq 0$. Since $\nabla^2 f(\Phi\bar{x}) \succ 0$, we have $\Phi w = 0$. This together with (4.33) tells us that $w \in \text{Ker}\,\Phi \cap \mathcal{C}_J(x_0) = \{0\}$ by (4.31). Thus inequality (4.34) holds and $\bar{x}$ is a strong solution of problem (4.27).

(iii) Suppose that $\bar{x}$ is a strong solution of problem (4.27). There exist constants $\kappa, \gamma > 0$ such that

$$f(\Phi x) + \mu J(x) \geq f(\Phi \bar{x}) + \mu J(\bar{x}) + \frac{\kappa}{2}\|x - \bar{x}\|^2 \quad \text{for all} \quad x \in \mathbb{B}_\gamma(\bar{x}).$$

For any $x \in \mathbb{B}_\gamma(\bar{x}) \cap \Phi^{-1}(\Phi\bar{x})$, we obtain that

$$J(x) \geq J(\bar{x}) + \frac{\kappa}{2\mu}\|x - \bar{x}\|^2,$$

which means that $\bar{x}$ is also a strong solution of (1.2) as claimed.

□

For the special case (1.5) of (4.27), [15, Theorem 1] shows that if $\bar{x}$ is an optimal solution of (4.27) with $x_0 = \bar{x}$ and the Strong Null Space Property (4.22) holds at $\bar{x}$, $\bar{x}$ is the unique solution of (4.27). As Strong Null Space Property is a characterization for sharp minima to problem (1.2), Proposition 4.13 advances [15, Theorem 1] and [43, Theorem 3] with further information that $\bar{x}$ is a strong solution of (1.5).

Two natural questions arise from Proposition 4.13: are the converse statements of (ii)-(iii) true ? That is:

(Q.1) If $\bar{x}$ is a strong solution of (4.27), can it be a sharp solution of (1.2) with $x_0 = \bar{x}$?

(Q.2) If $\bar{x}$ is a strong solution of (1.2), can it be a strong solution of (4.27)?

For analysis ℓ_1 problems, i.e., $J_0(\cdot) = \|\cdot\|_1$, and more generally for J_0 the support function of any polyhedral convex compact C such that $0 \in \operatorname{ri} C$, we have positive answers for both questions.

Corollary 4.14 (Solution uniqueness to ℓ_1 problems)**.** *Let $\bar{x}$ be a minimizer of* (4.27) *with $J_0(\cdot) = \|\cdot\|_1$. Then the following are equivalent:*

(i) *$\bar{x}$ is the unique solution of problem* (1.2) *with $x_0 = \bar{x}$.*

(ii) *$\bar{x}$ is the sharp solution of problem* (1.2) *with $x_0 = \bar{x}$.*

(iii) *$\bar{x}$ is the unique solution of problem* (4.27)*.*

(iv) *$\bar{x}$ is the strong solution of problem* (4.27)*.*

Proof. [(i)⇔(iii)] and [(ii)⇒(iv)] are from Proposition 4.13. [(iv)⇒(iii)] is trivial. Finally, [(i)⇔(ii)] follows by combining Proposition 3.4 and Proposition 3.8(i) since J_0 is a polyhedral norm, in which case the descent and critical cones coincide at any x_0 (see Proposition 3.7). □

According to Theorem 4.6, conditions $\operatorname{Ker}\Phi \cap \operatorname{Ker} D_S^* = \{0\}$ and $\rho(e) < 1$ form a characterization for solution uniqueness to ℓ_1 problem. A similar result was established in [48, Theorem 1] for problems (1.2) and (1.5) with the extra assumption that Φ has full row-rank. Corollary 4.14 is more general and reveals that the unique solution of problem (4.27) is the strong solution.

The answer is negative for **(Q.1)** in general; see our Theorem 5.12 and Example 3.9. Regarding **(Q.2)**, we do have a positive answer for group-sparsity ; see Theorem 5.3 and Theorem 5.12. However, it is not true for the nuclear norm minimization problem as we now show.

Example 4.15 (Strong solutions of (1.2) are not those of (4.27))**.** *Let us consider the following nuclear norm minimization problem*

$$\min_{X\in\mathbb{R}^{2\times 2}} \Theta(X) \stackrel{\text{def}}{=} \frac{1}{2}\|Y-\Phi(X)\|^2 + \|X\|_*, \tag{4.35}$$

where $\|\cdot\|_*$ *stands for the nuclear norm of* X*, and* $\Phi:\mathbb{R}^{2\times 2}\to\mathbb{R}^2$ *is the diagonal operator, i.e.,* $\Phi(X) = (X_{11},X_{22})^\top$ *and* $Y=(2,1)^\top$*. This is a special case of (4.27) with* $f(\cdot)=\frac{1}{2}\|Y-\cdot\|^2$*. Let* $\overline{X}=\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}$*. We have*

$$\Phi^*\nabla f(\Phi(\overline{X})) = \begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix} \text{ and } \partial\|\overline{X}\|_* = \left\{\begin{pmatrix}1 & 0\\ 0 & \alpha\end{pmatrix} \mid \alpha\in[-1,1]\right\}.$$

Thus $0\in\partial\Theta(\overline{X})$*, i.e.,* $\overline{X}$ *is a solution of (4.35). By Proposition 4.13, it is also a solution of*

$$\min_{X\in\mathbb{R}^{2\times 2}} \|X\|_* \text{ and } \Phi(X)=(1,0)^\top. \tag{4.36}$$

For any 2×2 *matrix* X*, let* σ_1,σ_2 *be its singular values. We have*

$$\|X\|_* = \sigma_1+\sigma_2 = \sqrt{\sigma_1^2+\sigma_2^2+2\sigma_1\sigma_2} = \sqrt{\|X\|_F^2+2|\det(X)|}. \tag{4.37}$$

The feasible set of (4.36) consists of matrices of the form $X=\begin{pmatrix}1 & a\\ b & 0\end{pmatrix}$*, and we obtain from (4.37) that*

$$\|X\|_* = \sqrt{1+(|a|+|b|)^2}.$$

Thus $\overline{X}$ *is the unique solution of (4.36) as well as (4.35) due to Proposition 4.13 again. Next we claim that* $\overline{X}$ *is the strong solution of (4.36), which means there exist* $\kappa,\delta>0$ *such that*

$$\|X\|_* - \|\overline{X}\|_* \ge \frac{\kappa}{2}\|X-\overline{X}\|_F^2 \quad \text{for all} \quad X=\begin{pmatrix}1 & a\\ b & 0\end{pmatrix} \quad \text{with} \quad \|X-\overline{X}\|_F = \sqrt{a^2+b^2}\le\delta. \tag{4.38}$$

Indeed, we have

$$\begin{aligned}\|X\|_* - \|\overline{X}\|_* &= \sqrt{1+(|a|+|b|)^2}-1\\ &= \frac{a^2+b^2+2|ab|}{\sqrt{1+a^2+b^2+2|ab|}+1}\\ &\ge \frac{a^2+b^2}{\sqrt{1+2\delta^2}+1}\\ &= \frac{1}{\sqrt{1+2\delta^2}+1}\|X-\bar{X}\|_F^2.\end{aligned}$$

This certainly verifies (4.38). Next we claim that $\overline{X}$ *is not a strong solution of (4.35). Pick* $X_\varepsilon \stackrel{\text{def}}{=} \begin{pmatrix}1+\varepsilon^2 & \varepsilon\\ \varepsilon & \varepsilon^2\end{pmatrix}$ *with* $\varepsilon>0$ *sufficiently small, observe that* $\|X-\overline{X}\|_F^2 = 2(\varepsilon^2+\varepsilon^4)$*. It follows from (4.37) that*

$$\begin{aligned}\Theta(X_\varepsilon)-\Theta(\overline{X}) &= \frac{1}{2}\left\|\begin{pmatrix}\varepsilon^2-1\\ \varepsilon^2-1\end{pmatrix}\right\|^2 + \left\|\begin{pmatrix}1+\varepsilon^2 & \varepsilon\\ \varepsilon & \varepsilon^2\end{pmatrix}\right\|_* - 2\\ &= (\varepsilon^2-1)^2 + \sqrt{((1+\varepsilon^2)^2+2\varepsilon^2+\varepsilon^4)+2\varepsilon^4}-2\\ &= (\varepsilon^2-1)^2+(1+2\varepsilon^2)-2\\ &= \varepsilon^4.\end{aligned}$$

Therefore, $\overline{X}$ cannot be a strong solution of (4.35), which also implies that $\overline{X}$ is neither a sharp solution of (4.35) nor to (4.36) due to Proposition 4.13.

5 Characterizations of unique/strong solutions for group-sparsity

In this section, we study the following particular case of (1.2), where J_0 is the ℓ_1/ℓ_2 norm that promotes group sparsity [32, 36, 46]

$$\min_{x\in\mathbb{R}^n} \quad \|D^*x\|_{\ell_1/\ell_2} \quad \text{subject to} \quad \Phi x = \Phi x_0. \tag{5.1}$$

Following the notation in [8], we suppose that $\mathbb{R}^p$ is decomposed into q groups by

$$\mathbb{R}^p = \bigoplus_{g=1}^{q} V_g, \tag{5.2}$$

where each V_g is a subspace of $\mathbb{R}^p$ with the same dimension G. For any $u \in \mathbb{R}^p$, we write $u = \sum_{g=1}^{q} u_g$ with $u_g \in V_g$ being the vector in group V_g. The ℓ_1/ℓ_2 norm in $\mathbb{R}^p$ is defined by

$$\|u\|_{\ell_1/\ell_2} = \sum_{g=1}^{q} \|u_g\|. \tag{5.3}$$

Its dual is the ℓ_∞/ℓ_2 norm:

$$\|u\|_{\ell_\infty/\ell_2} = \max_{1\le g\le q} \|u_g\|. \tag{5.4}$$

With $\bar{u} \stackrel{\text{def}}{=} D^*x_0$, define $I \stackrel{\text{def}}{=} \{g \in \{1,\ldots,q\}|\ \bar{u}_g \neq 0\}$, the index set of active groups of $\bar{u}$, and $K \stackrel{\text{def}}{=} \{1,\ldots,q\} \setminus I$, the index set of nonactive groups of $\bar{u}$. Note that

$$\partial\|\bar{u}\|_{\ell_1/\ell_2} \stackrel{\text{def}}{=} \left\{v \in \mathbb{R}^p|\ v_g = \frac{\bar{u}_g}{\|\bar{u}_g\|_2} \text{ for } g \in I \text{ and } \|v_g\| \le 1 \text{ for } g \in K\right\}. \tag{5.5}$$

Thus ℓ_1/ℓ_2 norm is decomposable at $\bar{u}$ as in Remark 4.5, where $T = \bigoplus_{g\in I} V_g$, $S = T^\perp$, and

$$e = \sum_{g\in I} \frac{\bar{u}_g}{\|\bar{u}_g\|}. \tag{5.6}$$

5.1 Descent cone of group sparsity

Sharp minima at x_0 for problem (5.1) is studied in our previous section. However, unlike the case of ℓ_1 problem, a unique solution of problem (5.1) may be not sharp; see Example 3.9. To characterize the solution uniqueness to group sparsity problem (5.1), we compute the descent cone $\mathcal{D}_J(x_0)$ in (3.8) as follows.

Theorem 5.1 (Descent cone to ℓ_1/ℓ_2 problem and geometric characterization for solution uniqueness)**.** *The descent cone at x_0 to problem (5.1) is given by*

$$\mathcal{D}_J(x_0) = (\mathcal{E} \cap \operatorname{bd} \mathcal{C}_J(x_0)) \cup (\operatorname{int} \mathcal{C}_J(x_0)), \tag{5.7}$$

where

$$\mathcal{E} \stackrel{\text{def}}{=} \left\{w \in \mathbb{R}^n \middle|\ D_T^* w \in \operatorname{Im}\left\{(D_T^* x_0)_g \middle|\ g \in I\right\}\right\}, \tag{5.8}$$

and $\operatorname{bd}\mathcal{C}_J(x_0)$ *stands for the boundary of the critical cone* $\mathcal{C}_J(x_0)$ *defined in* (3.10)*, i.e.,*

$$\operatorname{bd}\mathcal{C}_J(x_0) = \{w \in \mathbb{R}^p |\ \langle D_T e, w\rangle + \|D_S^* w\|_{\ell_1/\ell_2} = 0\}. \tag{5.9}$$

Consequently, x_0 *is a unique solution of* (5.1) *if and only if*

$$\operatorname{Ker}\Phi \cap \Big((\mathcal{E} \cap \operatorname{bd}\mathcal{C}_J(x_0)) \cup (\operatorname{int}\mathcal{C}_J(x_0))\Big) = \{0\}. \tag{5.10}$$

Proof. Let us start by verifying the inclusion "$\supset$" in (5.7). Recall that $J(x) = \|D^* x\|_{\ell_1/\ell_2}$. For any $w \in \operatorname{int}(\mathcal{C}_J(x_0))$, we get from (3.10) that

$$dJ(x_0)(w) = \lim_{t\downarrow 0} \frac{J(x_0 + tw) - J(x_0)}{t} < 0.$$

Hence there is some $t_0 > 0$ such that $J(x_0 + t_0 w) < J(x_0)$. This ensures $w \in \mathcal{D}_J(x_0)$. For any $w \in \mathcal{E} \cap \operatorname{bd}\mathcal{C}_J(x_0)$, we represent $D_T^* w = \sum_{g\in I} \lambda_g (D^* x_0)_g$ with $\lambda_g \in \mathbb{R}$, $g \in I$ and

$$0 = \langle D_T e, w\rangle + \|D_S^* w\|_{\ell_1/\ell_2} = \langle e, D_T^* w\rangle + \|D_S^* w\|_{\ell_1/\ell_2} = \sum_{g\in I} \lambda_g \|(D^* x_0)_g\| + \|D_S^* w\|_{\ell_1/\ell_2}. \tag{5.11}$$

Define $I_- \stackrel{\text{def}}{=} \{g \in I |\ \lambda_g < 0\}$. If $I_- = \emptyset$, we get from (5.11) that $D_S^* w = 0$ and $\lambda_g = 0$ for all $g \in I$, which implies that $D_T^* w = 0$ and thus $D^* w = 0$. Therefore, we get

$$J(x_0 + w) = J_0(D^*(x_0 + w)) = J_0(D^* x_0) = J(x_0),$$

which tells us that $w \in \mathcal{D}_J(x_0)$. If $I_- \neq \emptyset$, define $t_1 \stackrel{\text{def}}{=} \min\left\{-\frac{1}{\lambda_g} \middle|\middle|\ g \in I_-\right\} > 0$, we obtain from (5.11) that

$$\begin{aligned}
J(x_0 + t_1 w) - J(x_0) &= \|D^* x_0 + t_1 D^* w\|_{\ell_1/\ell_2} - \|D^* x_0\|_{\ell_1/\ell_2} \\
&= \sum_{g\in I} |1 + t_1 \lambda_g| \|(D^* x_0)_g\| + t_1 \|D_S^* w\|_{\ell_1/\ell_2} - \|D^* x_0\|_{\ell_1/\ell_2} \\
&= \sum_{g\in I} (1 + t_1 \lambda_g) \|(D^* x_0)_g\| + t_1 \|D_S^* w\|_{\ell_1/\ell_2} - \|D^* x_0\|_{\ell_1/\ell_2} \\
&= t_1 \left(\sum_{g\in I} \lambda_g \|(D^* x_0)_g\| + \|D_S^* w\|_{\ell_1/\ell_2}\right) + \sum_{g\in I} \lambda_g \|(D^* x_0)_g\| - \|D^* x_0\|_{\ell_1/\ell_2} \\
&= 0,
\end{aligned}$$

where the third equality is from the choice of t_1. In both cases of I_-, $w \in \mathcal{D}_J(x_0)$.

To justify the reverse inclusion "$\subset$" in (5.7), take any $w \in \mathcal{D}_J(x_0)$ and let $t > 0$ such that $\|D^* x_0 + tD^* w\|_{\ell_1/\ell_2} \le \|D^* x_0\|_{\ell_1/\ell_2}$. For any $\alpha \in (0, 1)$, we have

$$\|D^* x_0\|_{\ell_1/\ell_2} \ge \alpha \|D^* x_0 + tD^* w\|_{\ell_1/\ell_2} + (1 - \alpha)\|D^* x_0\|_{\ell_1/\ell_2} \ge \|D^* x_0 + \alpha t D^* w\|_{\ell_1/\ell_2}. \tag{5.12}$$

Choose $\alpha > 0$ sufficiently small such that

$$\|(D^* x_0)_g + \alpha t (D^* w)_g\| > 0 \quad \text{for all} \quad g \in I.$$

Since $w \in \mathcal{D}_J(x_0) \subset \mathcal{C}_J(x_0)$ by (3.11), it suffices to show that if $w \in \operatorname{bd} \mathcal{C}_J(x_0)$ then $w \in \mathcal{E}$. Suppose that $w \in \operatorname{bd} \mathcal{C}_J(x_0)$, we have

$$0 = \langle D_T e, w \rangle + \|D_S^* w\|_{\ell_1/\ell_2} = \langle e, D_T^* w \rangle + \|D_S^* w\|_{\ell_1/\ell_2}.$$

It follows from the latter and (5.12) that

$$\begin{aligned} 0 &\geq \|D^* x_0 + \alpha t D^* w\|_{\ell_1/\ell_2} - \|D^* x_0\|_{\ell_1/\ell_2} \\ &= \sum_{g \in I} \|(D^* x_0)_g + \alpha t (D^* w)_g\| + \alpha t \sum_{g \in K} \|(D^* w)_g\| - \|D^* x_0\|_{\ell_1/\ell_2} \\ &= \sum_{g \in I} \Big(\|(D^* x_0)_g + \alpha t (D^* w)_g\| - \|(D^* x_0)_g\| - \alpha t \left\langle e_g, (D^* w)_g \right\rangle \Big). \end{aligned} \tag{5.13}$$

Since $e_g = \partial \|(D^* x_0)_g\|$, each $\|(D^* x_0)_g + \alpha t (D^* w)_g\| - \|(D^* x_0)\|_g - \alpha t \left\langle e_g, (D^* w)_g \right\rangle \geq 0$ for $g \in I$. This together with (5.13) tells us that

$$0 = \|(D^* x_0)_g + \alpha t (D^* w)_g\| - \|(D^* x_0)_g\| - \alpha t \left\langle e_g, (D^* w)_g \right\rangle \quad \text{for all} \quad g \in I.$$

Hence, we have

$$0 = \|(D^* x_0)_g\| - \left\langle e_g, (D^* x_0)_g \right\rangle = \|(D^* x_0)_g + \alpha t (D^* w)_g\| - \left\langle e_g, (D^* x_0)_g + \alpha t (D^* w)_g \right\rangle.$$

As $\|e_g\| = 1$ and $(D^* x_0)_g + \alpha t (D^* w)_g \neq 0$, the latter equality holds when

$$(D^* x_0)_g + \alpha t (D^* w)_g = \delta_g e_g$$

for some $\delta_g > 0$, $g \in I$. It follows that

$$(D^* w)_g = \frac{1}{\alpha t} (\delta_g - \|(D^* x_0)_g\|) e_g$$

for any $g \in I$, which ensures that $w \in \mathcal{E}$ and verifies the equality (5.7).

The characterization for solution uniqueness at x_0 in (5.10) follows directly from Proposition 3.4 and (5.7). □

As observed in Example 3.9, the descent cone $\mathcal{D}_J(x_0)$ is different from the critical cone $\mathcal{C}_J(x_0)$. However, the closure of $\mathcal{D}_J(x_0)$ is $\mathcal{C}_J(x_0)$ thanks to Proposition 3.7. The decomposition (5.7) not only provides an explicit computation of $\mathcal{D}_J(x_0)$ via first-order information, but also identifies the key set $\mathcal{E}$ and indicates that the major geometric difference between these two cones lies in the boundary of $\mathcal{C}_J(x_0)$, i.e.,

$$\mathcal{D}_J(x_0) \setminus \operatorname{int} \mathcal{D}_J(x_0) = \mathcal{E} \cap \operatorname{bd} \mathcal{C}_J(x_0) \subset \operatorname{bd} \mathcal{C}_J(x_0).$$

This result is crucial in proving our Theorem 5.3 below, which shows the equivalence between unique solution and strong solution for problem (5.1) while providing a more computable characterization for solution uniqueness.

5.2 Unique vs strong solutions

We show next that a unique solution of (1.2) is indeed a strong solution. The proof is based on the second-order analysis in Lemma 2.4. We need the computation of the second subderivative for the function Ψ defined in (3.3).

Lemma 5.2 (Second subderivative to ℓ_1/ℓ_2 norm)**.** *Suppose that x_0 is a minimizer of problem (5.1). Then we have*

$$\operatorname{dom} d^2\Psi(x_0|0) = \operatorname{Ker}\Phi \cap \operatorname{bd}\mathcal{C}_J(x_0) \tag{5.14}$$

and

$$d^2\Psi(x_0|0)(w) = \sum_{g\in I} \frac{\|(D^*w)_g\|^2\|(D^*x_0)_g\|^2 - \langle (D^*x_0)_g, (D^*w)_g\rangle^2}{\|(D^*x_0)_g\|^3} \quad \textit{for} \quad w \in \operatorname{dom} d^2\Psi(x_0|0). \tag{5.15}$$

Proof. Since x_0 is an optimal solution of (5.1), we have $0 \in \partial\Psi(x_0)$. With $C = \Phi^{-1}(\Phi x_0)$, we have

$$\begin{aligned} d^2\Psi(x_0|0)(w) &= \liminf_{t\downarrow 0,\, w'\to w} \frac{\|D^*(x_0+tw')\|_{\ell_1/\ell_2} + \iota_C(x_0+tw') - \|D^*x_0\|_{\ell_1/\ell_2} - \iota_C(x_0) - t\langle 0, w\rangle}{\frac{1}{2}t^2} \\ &= \liminf_{t\downarrow 0,\, w'\to w} \left(\frac{\|D^*(x_0+tw')\|_{\ell_1/\ell_2} - \|D^*x_0\|_{\ell_1/\ell_2}}{\frac{1}{2}t^2} + \iota_{\operatorname{Ker}\Phi}(w') \right). \end{aligned}$$

Closedness of $\operatorname{Ker}\Phi$ implies that $\operatorname{dom} d^2\Psi(x_0|0) \subset \operatorname{Ker}\Phi$. Since $D_S^* x_0 = 0$, we have

$$\begin{aligned} &\liminf_{t\downarrow 0,\, w'\overset{\operatorname{Ker}\Phi}{\to} w} \frac{\|D^*(x_0+tw')\|_{\ell_1/\ell_2} - \|D^*x_0\|_{\ell_1/\ell_2}}{\frac{1}{2}t^2} \\ &= \liminf_{t\downarrow 0,\, w'\overset{\operatorname{Ker}\Phi}{\to} w} \left(\frac{\|D_T^*(x_0+tw')\|_{\ell_1/\ell_2} - \|D_T^*x_0\|_{\ell_1/\ell_2} - t\langle De, w'\rangle}{\frac{1}{2}t^2} + \frac{\langle De, w'\rangle + \|D_S^*w'\|_{\ell_1/\ell_2}}{\frac{1}{2}t} \right). \end{aligned} \tag{5.16}$$

Note that the Euclidean norm $\|u\|$ is twice differentiable at $u \neq 0$ with

$$\nabla\|u\| = \frac{u}{\|u\|} \quad \text{and} \quad \nabla^2\|u\| = \frac{1}{\|u\|}\mathbb{I} - \frac{1}{\|u\|^3}uu^*$$

This together with (2.16) tells us that

$$\begin{aligned} &\liminf_{t\downarrow 0,\, w'\overset{\operatorname{Ker}\Phi}{\to} w} \frac{\|D_T^*(x_0+tw')\|_{\ell_1/\ell_2} - \|D_T^*x_0\|_{\ell_1/\ell_2} - t\langle De, w'\rangle}{\frac{1}{2}t^2} \\ &= \liminf_{t\downarrow 0,\, w'\overset{\operatorname{Ker}\Phi}{\to} w} \sum_{g\in I} \left(\frac{\|(D^*(x_0+tw'))_g\| - \|(D^*x_0)_g\| - t\langle e_g, D^*w'\rangle}{\frac{1}{2}t^2} \right) \\ &= \sum_{g\in I} \left\langle (D^*w)_g, \frac{(D^*w)_g}{\|(D^*x_0)_g\|} - \frac{1}{\|(D^*x_0)_g\|^3}(D^*x_0)_g(D^*x_0)_g^*(D^*w)_g \right\rangle \\ &= \sum_{g\in I} \frac{\|(D^*w)_g\|^2\|(D^*x_0)_g\|^2 - \langle (D^*x_0)_g, (D^*w)_g\rangle^2}{\|(D^*x_0)_g\|^3}. \end{aligned} \tag{5.17}$$

Since x_0 is an optimal solution of (5.1), there exists $z \in \mathbb{R}^p$ with $P_T z = e$, $\|P_S z\|_{\ell_\infty/\ell_2} \le 1$ and $Dz \in \operatorname{Im}\Phi^*$. For any $w' \in \operatorname{Ker}\Phi$, we have

$$\langle De, w'\rangle = \langle D_T z, w'\rangle = -\langle D_S z, w'\rangle = -\langle z, D_S^* w'\rangle.$$

It follows that $\langle De, w'\rangle + \|D_S^* w'\|_{\ell_1/\ell_2} \geq 0$. Hence we get

$$\liminf_{t\downarrow 0,\, w' \overset{\mathrm{Ker}\,\Phi}{\to} w} \frac{\langle De, w'\rangle + \|D_S^* w'\|_{\ell_1/\ell_2}}{\frac{1}{2}t} = \begin{cases} 0 & \text{if } \langle De, w\rangle + \|D_S^* w\|_{\ell_1/\ell_2} = 0, w \in \mathrm{Ker}\,\Phi, \\ \infty & \text{otherwise.} \end{cases}$$

Since the "liminf" in (5.17) indeed becomes "lim", the latter together with (5.17) and (5.16) verifies (5.14) and (5.15). □

This calculation allows us to establish the main result in this section, which gives a quantitative characterization for unique/strong solutions to ℓ_1/ℓ_2 problem (5.1).

Theorem 5.3 (Characterizations for unique/strong solutions to ℓ_1/ℓ_2 problems)**.** *The following assertions are equivalent:*

(i) *x_0 is a unique solution of problem* (5.1)*.*

(ii) *x_0 is a strong solution of problem* (5.1)*.*

(iii) *x_0 is a solution of* (5.1)*, $\mathrm{Ker}\,\Phi \cap \mathcal{E} \cap \mathrm{Ker}\,D_S^* = \{0\}$, and*

$$\zeta(e) \overset{\text{def}}{=} \min_{u\in \mathrm{Ker}\, MD_S} \|(MD_S)^\dagger MD_T e - u\|_{\ell_\infty/\ell_2} < 1, \tag{5.18}$$

where M^* is a matrix forming a basis matrix to $\mathrm{Ker}\,\Phi \cap \mathcal{E}$.

Proof. We first claim that x_0 is a solution of (5.1) if and only if

$$\mathrm{Ker}\,\Phi \cap \mathrm{int}\,(\mathcal{C}_J(x_0)) = \emptyset. \tag{5.19}$$

Indeed, x_0 is a solution of (5.1) if and only if $d\Psi(x_0)(w) \geq 0$ for all $w \in \mathbb{R}^n$. Due to the computation of $d\Psi(\bar{x})(w)$ in (4.15), $d\Psi(\bar{x})(w) \geq 0$ means

$$\langle De, w\rangle + \|D_S^* w\|_{\ell_1/\ell_2} \geq 0 \quad \text{for all} \quad w \in \mathrm{Ker}\,\Phi, \tag{5.20}$$

which is equivalent to (5.19).

Next let us verify the equivalence between (i) and (ii). By Theorem 5.1, it suffices to show that condition (5.10) implies (ii). Note from (5.14) and (5.15) that if $d^2\Psi(x_0|0)(w) \leq 0$ then $w \in \mathrm{Ker}\,\Phi \cap \mathrm{bd}\,\mathcal{C}_J(x_0)$ and $w \in \mathcal{E}$. Since $\mathrm{Ker}\,\Phi \cap (\mathcal{E} \cap \mathrm{bd}\,\mathcal{C}_J(x_0)) = \{0\}$ by (5.10), we have $d^2\Psi(x_0|0)(w) > 0$ for all $w \neq 0$. It follows from Lemma 2.4 that x_0 is a strong solution. This verifies the equivalence between (i) and (ii).

To justify the equivalence between (i) and (iii), by (5.19), we only need to show that condition (recall (5.10))

$$\mathrm{Ker}\,\Phi \cap \mathcal{E} \cap \mathrm{bd}\,\mathcal{C}_J(x_0) = \{0\} \tag{5.21}$$

is equivalent to the combination of $\mathrm{Ker}\,\Phi \cap \mathcal{E} \cap \mathrm{Ker}\,D_S^* = \{0\}$ and (5.18) provided that x_0 is a solution of (5.1). According to (5.20), (5.21) is equivalent to the condition that there exists some $c > 0$ such that

$$k(w) \overset{\text{def}}{=} \langle De, w\rangle + \|D_S^* w\|_{\ell_1/\ell_2} \geq c\|w\| \quad \text{for all} \quad w \in \mathrm{Ker}\,\Phi \cap \mathcal{E}. \tag{5.22}$$

Since x_0 is a solution of (5.1), there exists $v \in \mathbb{R}^p$ such that $v \in \partial\|D^* x_0\|_{\ell_1/\ell_2}$ and $Dv \in \mathrm{Im}\,\Phi^*$. It follows that $D(e + P_S v) \in \mathrm{Ker}\,M = \mathrm{Im}\,\Phi^* + \mathcal{E}^\perp$, we have $MD_S v = -MDe$. It follows from (2.2) that

$$MD_S(MD_S)^\dagger MD_T e = MD_T e.$$

For any $w = M^*u \in \operatorname{Ker}\Phi \cap C$ it follows that

$$\begin{aligned} k(w) &= k(M^*u) = \langle De, M^*u\rangle + \|D_S^*M^*u\|_{\ell_1/\ell_2} \\ &= \langle MD_T e, u\rangle + \|D_S^* w\|_{\ell_1/\ell_2} \\ &= \langle (MD_S)^\dagger MD_T e, D_S^* M^* u\rangle + \|D_S^*M^*u\|_{\ell_1/\ell_2}. \end{aligned} \tag{5.23}$$

Mimicking the proof of Theorem 4.6 by replacing N there by M and $\operatorname{Ker}\Phi$ by its subspace $\operatorname{Ker}\Phi \cap \mathcal{E}$, the inequality $k(w) \ge c\|w\|$ in (5.22) is equivalent to (iii). The proof is complete. □

It is worth noting that condition

$$\operatorname{Ker}\Phi \cap \mathcal{E} \cap \operatorname{Ker} D_S^* = \{0\} \tag{5.24}$$

in part (iii) is strictly weaker than the Restricted Injectivity in Theorem 4.6; see Example 3.9. It means that Φ is injective on the subspace $\mathcal{E} \cap \operatorname{Ker} D_S^*$. We refer (5.24) as *Strong Restricted Injectivity* condition. Moreover, we call the constant $\zeta(e)$ *Strong Source Coefficient*, while the condition $\zeta(e) < 1$ is refered as *Analysis Nondegenerate Source Condition* for solution uniqueness to ℓ_1/ℓ_2 problem (1.2).

Remark 5.4 (Checking the Strong Restricted Injectivity and the Analysis Nondegenerate Source Condition)**.** *Set the matrix* $S \overset{\text{def}}{=} ((D_T^* x_0)_g)_{g\in I}$ *to be an* $p \times |I|$ *matrix, where* $|I|$ *is the cardinality of* I*. Observe from* (5.15) *that* $w \in \mathcal{E}$ *if and only if the following system*

$$D_T^* w = S\lambda \textit{ is consistent with } \lambda \in \mathbb{R}^{|I|}.$$

Since S *is injective, we have* $\lambda = S^\dagger D_T^* w = (S^*S)^{-1} S^* D_T^* w$ *due to* (2.2)*. It follows that* $w \in \mathcal{E}$ *if and only if*

$$S(S^*S)^{-1}S^*D_T^*w - D_T^*w = 0 \tag{5.25}$$

Note that $S^*S = \operatorname{diag}(\|(D_T^*x_0)_g\|^2)_{b\in I}$ *is an* $|I| \times |I|$ *diagonal matrix. Representing* S *in terms of groups, we have*

$$S(S^*S)^{-1}S^* = \operatorname{diag}\left(\sum_{g\in I} \|(D_T^*x_0)_g\|^{-2}\delta_g\right) SS^*,$$

where $\delta_g = (0,\ldots,0,\underbrace{1,1,\ldots,1}_{\textit{in } V_g},0,\ldots,0)$ *is the unit vector in* V_g*. This together with* (5.25) *tells us that* $\mathcal{E}$ *is the kernel of the following matrix*

$$Q = \operatorname{diag}\left(\sum_{g\in I} \|(D_T^*x_0)_g\|^2\delta_g\right) D_T^* - SS^*D_T^* \quad \textit{with} \quad S = ((D_T^*x_0)_g)_{g\in I}. \tag{5.26}$$

The Strong Restricted Injectivity (5.24) *is equivalent to* $\operatorname{Ker}\begin{pmatrix}\Phi \\ Q\end{pmatrix} \cap \operatorname{Ker} D_S^* = \{0\}$.

Furthermore, M^* *forms a basis matrix of* $\operatorname{Ker}\begin{pmatrix}\Phi \\ Q\end{pmatrix}$*, which is found from the SVD of* $\begin{pmatrix}\Phi \\ Q\end{pmatrix}$*. Similarly to* (4.11)*,* $\zeta(e)$ *is the optimal solution of*

$$\min \|z\|_{\ell_2/\ell_\infty} \quad \textit{such that} \quad MDz = -MDe \;\; \textit{and} \;\; z \in \bigoplus_{g\in I} V_g. \tag{5.27}$$

So $\zeta^2(e)$ is the optimal value to the following convex optimization problem

$$\min_{t\geq 0, z} t \quad \text{such that} \quad MDz = -MDe, \quad \|z_g\|^2 - t \leq 0, \quad g \in I \quad \text{and} \quad z \in \bigoplus_{g\in I} V_g \tag{5.28}$$

with $|I|G + 1$ variables, which can be solved by available packages such as `cvxopt`*; see Section 6 for further discussion.*

Next we show that $\zeta(e) \leq \rho(e)$ when x_0 is an optimal solution of (5.1).

Proposition 5.5 (Comparison between $\rho(e)$ and $\zeta(e)$)**.** *Suppose that x_0 is an optimal solution of* (5.1)*. Then we have $\zeta(e) \leq \rho(e) \leq 1$.*

Proof. It follows from (4.6) that

$$ND_S(ND_S)^\dagger ND_T e = ND_T e \ \text{ and } \ \rho(e) \leq 1.$$

For any $w \in \operatorname{Ker}\Phi \cap \mathcal{E} = \operatorname{Im} M^* \subset \operatorname{Ker}\Phi = \operatorname{Im} N^*$, it is similar to (5.23) that

$$\langle De, w\rangle = \langle (MD_S)^\dagger MD_T e, D_S^* w\rangle = \langle (ND_S)^\dagger ND_T e, D_S^* w\rangle$$

Note from the definition in (5.18) that

$$\begin{aligned} \zeta(e) &= \sup_{\|v\|\leq 1} \langle -(MD_S)^\dagger MD_T e, v\rangle - \iota_{\operatorname{Im} D_S^* M^*}(v) \\ &= \sup_{\|v\|\leq 1} \langle -(ND_S)^\dagger ND_T e, v\rangle - \iota_{\operatorname{Im} D_S^* M^*}(v) \\ &= \min_{u\in \operatorname{Ker} MD_S} \|(ND_S)^\dagger ND_T e - u\|_{\ell_\infty/\ell_2}. \end{aligned} \tag{5.29}$$

Since $\operatorname{Ker} ND_S \subset \operatorname{Ker} MD_S$, we have $\zeta(e) \leq \rho(e)$. □

Although $\zeta(e)$ could be computed by involving N via (5.29), the format in (5.18) is more preferable. This is due to the fact that the Moore-Penrose inverse $MD_S^\dagger$ has a closed form as $(MD_S)^*(MD_S(MD_S)^*)^{-1}$ when the Strong Restricted Injectivity (5.24) holds, i.e., $D_S^* M^*$ is injective. In general, $\rho(e)$ is strictly smaller $\zeta(e)$. This fact is obtained through numerical experiments in our Section 6.

By replacing $\zeta(e)$ by $\rho(e)$ in Theorem 5.3, we do not need to assume x_0 to be an optimal solution, but we only have a sufficient condition for solution of uniqueness.

Corollary 5.6 (Sufficient condition for solution uniqueness to ℓ_1/ℓ_2 problem)**.** *x_0 is the unique solution of problem* (5.1) *provided that $\rho(e) < 1$ and both conditions* (4.6) *and* (5.24) *are satisfied.*

Proof. It follows directly from Proposition 4.4, Theorem 5.3, and Proposition 5.5. □

An simple upper bound for $\zeta(e)$ is

$$\zeta(e) \leq \gamma(e) \stackrel{\text{def}}{=} \|(MD_S)^\dagger MD_T e\|_{\ell_\infty/\ell_2}, \tag{5.30}$$

which is also used in Section 6 to check solution uniqueness. The inequality is indeed strict. The following result is straightforward from Theorem 5.3.

Corollary 5.7 (Sufficient condition for solution uniqueness to ℓ_1/ℓ_2 problem)**.** *Suppose that x_0 is an optimal solution of* (5.1)*. Then the Strong Restricted Injectivity* (5.24) *at x_0 and $\gamma(e) < 1$ are sufficient for solution uniqueness at x_0. They become necessary conditions provided that MD_S is injective.*

In the spirit of Proposition 5.5, it is possible that x_0 is a unique solution of (5.1) but $\rho(e) = 1$. It means that the Nondegenerate Source Condition may not happen. However, the following result provides characterizations for solution uniqueness whose statement closely relates to the Nondegenerate Source Condition.

Corollary 5.8 (Characterization for solution uniqueness to problem (5.1))**.** *The following are equivalent:*

(i) *x_0 is the unique solution of problem* (5.1).

(ii) *With an arbitrary $v \in \mathbb{R}^p$ satisfying $Dv \in \operatorname{Im}\Phi^*$, $P_T v = e$, and $\|P_S v\|_{\ell_\infty/\ell_2} \le 1$, the following system has only trivial solution*

$$\Phi w = 0, \quad w \in \mathcal{E}, \ \text{ and } \ D_S^* w \in \operatorname{cone}\{v_g | g \in L\} \times \{0_H\},$$

where $L \overset{\text{def}}{=} \{g \in K |\ \|v_g\| = 1\}$ and $H \overset{\text{def}}{=} K \setminus L$.

Proof. Pick an arbitrary $v \in \mathbb{R}^p$ satisfying $Dv \in \operatorname{Im}\Phi^*$, $P_T v = e$, and $\|P_S v\|_{\ell_\infty/\ell_2} \le 1$. Such an v always exists as when x_0 is a solution of (5.1), i.e., $\operatorname{Ker}\Phi \cap \operatorname{int}\mathcal{C}_J(x_0) = \emptyset$. We claim that

$$\operatorname{Ker}\Phi \cap \operatorname{bd}\mathcal{C}_J(x_0) = \{w \in \operatorname{Ker}\Phi |\ D_S^* w \in \operatorname{cone}\{v_g | g \in L\} \times \{0_H\}\} \tag{5.31}$$

with $L = \{g \in K |\ \|v_g\| = 1\}$ and $H = K \setminus L$. For any $w \in \operatorname{Ker}\Phi$, note that

$$\begin{aligned}
\langle De, w\rangle + \|D_S^* w\| &= \langle Dv, w\rangle - \langle DP_S v, w\rangle + \|D_S^* w\| \\
&= -\langle P_S v, D_S^* w\rangle + \|D_S^* w\|_{\ell_1/\ell_2} \\
&= \sum_{g\in L} \left(-\langle v_g, (D_T^* w)_g\rangle + \|(D_T^* w)_g\|\right) + \sum_{g\in H} \left(\langle v_g, (D_T^* w)_g\rangle + \|(D_T^* w)_g\|\right) \\
&= \sum_{g\in L} \left(-\langle v_g, (D_T^* w)_g\rangle + \|(D_T^* w)_g\|\right) + \sum_{g\in H} \|(D_T^* w)_g\| \ge 0,
\end{aligned}$$

where we used Cauchy-Schwartz inequality and that $\|v_g\| = 1$, for all $g \in L$. It follows that $w \in \operatorname{bd}\mathcal{C}_J(x_0)$ if and only if there exist $\lambda_g \ge 0$, $g \in L$ such that $(D_T^* w)_g = \lambda_g v_g$ for any $g \in L$ and $(D_T^* w)_g = 0$ for any $g \in H$, which verifies (5.31). The equivalence between (i) and (ii) follows from Theorem 5.1. ☐

As $\operatorname{Ker}\Phi \cap \mathcal{E} \cap \operatorname{bd}\mathcal{C}_J(x_0) \subset \operatorname{Ker}\Phi \cap \operatorname{Ker} D_{V_H}^*$ with $V_H \overset{\text{def}}{=} \bigoplus_{g\in H} V_g$, the following result is straightforward from Corollary 5.6.

Corollary 5.9 (Sufficient condition to the solution uniqueness to problem (5.1))**.** *x_0 is the unique solution of problem* (5.1) *provided that there exists some v such that $Dv \in \operatorname{Im}\Phi^*$, $P_T v = e$, $\|P_S v\|_{\ell_\infty/\ell_2} \le 1$, and that*

$$\operatorname{Ker}\Phi \cap \operatorname{Ker} D_{V_H}^* = \{0\} \tag{5.32}$$

with $K = \{g \in J |\ \|v_g\| = 1\}$, $H = J \setminus K$, and $V_H \overset{\text{def}}{=} \bigoplus_{g\in H} V_g$.

The sufficient condition in [19, Proposition 7.1] is a special case of Corollary 5.9 where D^* is the discrete gradient operator. Moreover, [24, Theorem 3.4] and [36, Theorem 3] even assume a stronger condition as they require $K = \emptyset$. Let us revisit Example 3.9. Pick any $v \in \partial\|x_0\|_{\ell_1,\ell_2} \cap \operatorname{Im}\Phi^*$, i.e., $v \in (0,1) \times [-1,1] \cap \operatorname{Im}\{(1,1,0),(1,0,-1)\}$. It follows that $v = (0,1,1)$ and thus $K = \{3\}$, $H = \emptyset$, which clearly implies (5.32).

According to Theorem 5.3, Theorem 3.12, and Proposition 3.1, solution uniqueness to group-sparsity problem (5.1) is equivalent to the robust recovery with rate $\mathcal{O}(\sqrt{\delta})$.

Corollary 5.10 (Robust recovery and solution uniqueness for group-sparsity problems)**.** *The following statements are equivalent:*

(i) x_0 *is a unique solution of* (5.1).

(ii) *For sufficiently small* $\delta > 0$*, any solution* x_δ *to problem* (1.4) *with* $J_0(\cdot) = \|\cdot\|_{\ell_1/\ell_2}$ *satisfies* $\|x_\delta - x_0\| \le \mathcal{O}(\sqrt{\delta})$ *whenever* $\|y - y_0\| \le \delta$.

(iii) *For any* $c_1 > 0$ *and sufficiently small* $\delta > 0$*, any solution* x_μ *to* (1.5) *with* $J_0(\cdot) = \|\cdot\|_{\ell_1/\ell_2}$ *satisfies* $\|x_\mu - x_0\| \le \mathcal{O}(\sqrt{\delta})$ *whenever* $\|y - y_0\| \le \delta$ *and* $\mu = c_1\delta$.

From Theorem 5.3, it is natural to raise the following question for other decomposable norm minimization problem.

(Q.3) If x_0 is a unique solution of (1.2), can it be a strong solution of (1.2)?

The answer to **(Q.3)** is affirmative for ℓ_1/ℓ_2 problem as in Theorem 5.3. For the ℓ_1 problem when $\|\cdot\|_{\mathcal{A}} = \|\cdot\|_1$, we also have the positive answer due to Corollary 4.14. However, it is not the case for the nuclear norm minimization problem. The following example modifies [4, Example 3.1] to prove that claim.

Example 5.11 (Difference between unique solution and strong solution of NNM)**.** *Consider the following optimization problem*

$$\min_{X\in\mathbb{R}^{2\times 2}} \|X\|_* \quad \text{such that} \quad \Phi(X) \stackrel{\text{def}}{=} \begin{pmatrix} X_{11} + X_{22} \\ X_{12} - X_{21} + X_{22} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{5.33}$$

For any $X = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *with* $a + d = 1$ *and* $b - c + d = 0$*, we obtain from* (4.37) *that*

$$\|X\|_* = \sqrt{a^2 + b^2 + c^2 + d^2 + 2|ad - bc|} \ge \sqrt{|a + d|^2 + |b - c|^2} \ge 1,$$

where the equality occurs when $\|X\|_* = 1$*,* $b - c = 0$*,* $a + d = 1$*, and* $b - c + d = 0$*, which means* $b = c$*,* $a = 1$*,* $d = 0$*, and* $\|X\|_* = \sqrt{1 + b^2 + c^2 + 2|bc|} = 1$*. So* $\overline{X} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ *is the unique solution of problem* (5.33)*. Choose* $X_\varepsilon = \begin{pmatrix} 1 - \varepsilon^{1.5} & \varepsilon - \varepsilon^{1.5} \\ \varepsilon & \varepsilon^{1.5} \end{pmatrix}$ *with* $\varepsilon > 0$ *sufficiently small and note that* X_ε *satisfies the equation in* (5.33)*. It follows that*

$$\begin{aligned} \|X_\varepsilon\|_* - \|\overline{X}\|_* &= \sqrt{(1-\varepsilon^{1.5})^2 + (\varepsilon - \varepsilon^{1.5})^2 + \varepsilon^2 + \varepsilon^3 + 2|(1-\varepsilon^{1.5})\varepsilon^{1.5} - (\varepsilon - \varepsilon^{1.5})\varepsilon|} - 1 \\ &= \sqrt{(1-\varepsilon^{1.5})^2 + (\varepsilon - \varepsilon^{1.5})^2 + \varepsilon^2 + \varepsilon^3 + 2(1-\varepsilon^{1.5})\varepsilon^{1.5} - 2(\varepsilon - \varepsilon^{1.5})\varepsilon} - 1 \\ &= \sqrt{1 + \varepsilon^3} - 1 = \mathcal{O}(\varepsilon^3). \end{aligned}$$

Moreover, $\|X_\varepsilon - \overline{X}\|_F^2 = \varepsilon^3 + (\varepsilon - \varepsilon^{1.5})^2 + \varepsilon^2 + \varepsilon^3 = O(\varepsilon^2)$*. This tells us that* $\overline{X}$ *is not a strong solution of* (5.33)*.*

5.3 Connections between unique/strong solutions in the noiseless case

Let us consider the particular case of problem (4.27), which reads

$$\min \Theta(x) \stackrel{\text{def}}{=} f(\Phi x) + \mu \|D^* x\|_{\ell_1/\ell_2} \tag{5.34}$$

with constant $\mu > 0$. In the following result, we show that a unique solution of (5.34) is also a strong solution. Consequently, all the characterizations for solution uniqueness to (5.1) in this section can be used to characterize unique/strong solutions to problem (5.34) due to Proposition 4.13. Moreover, this result gives an affirmative answer for **(Q.2)** in the previous section.

Theorem 5.12 (Characterization to solution uniqueness to ℓ_1/ℓ_2 regularized problem)**.** *Suppose that $\bar{x}$ is an optimal solution of problem* (5.34)*. The following are equivalent:*

(i) *$\bar{x}$ is the unique solution of* (5.34)*.*

(ii) *$\bar{x}$ is the strong solution of* (5.34)*.*

(iii) *$\bar{x}$ is the strong solution of* (5.1) *with $x_0 = \bar{x}$.*

Proof. (i) and (iii) are equivalent due to Proposition 4.13 and Theorem 5.3. [(ii)⇒(i)] is trivial. It suffices to verify [(iii)⇒(ii)]. Suppose that $\bar{x}$ is a strong solution of (5.1). By Theorem 5.3 and Theorem 5.1, $\operatorname{Ker}\Phi \cap \mathcal{E} \cap \operatorname{bd} \mathcal{C}_J(x_0) = \{0\}$. Since $\bar{x}$ is an optimal solution of (5.12), we have

$$0 \in \partial\Theta(\bar{x}) = \Phi^*\nabla f(\Phi\bar{x}) + D\partial\|D^*\bar{x}\|_{\ell_1/\ell_2} \subset D\partial\|D^*\bar{x}\|_{\ell_1/\ell_2} + \operatorname{Im}\Phi^*.$$

Thus there exists $\bar{z} \in \partial\|D^*\bar{x}\|_{\ell_1/\ell_2}$ such that $D\bar{z} = -\Phi^*\nabla f(\Phi\bar{x})$, $P_T\bar{z} = e$, and $\|P_S\bar{z}\|_{\ell_\infty/\ell_2} \le 1$, where e, T, and I are defined at the beginning at the section with $x_0 = \bar{x}$.

For any $w \in \mathbb{R}^n$, note further that

$$d^2\Theta(\bar{x}|0)(w) = \langle \Phi w, \nabla^2 f(\Phi\bar{x})\Phi w\rangle + d^2 J(\bar{x}|\, D\bar{z})(w) \tag{5.35}$$

with $J(x) = \|D^*x\|_{\ell_1/\ell_2}$. Since $D_S^*\bar{x} = 0$, we have

$$\begin{aligned} d^2 J(\bar{x}|\, D\bar{z})(w) &= \liminf_{t\downarrow 0,\, w'\to w} \frac{\|D^*(\bar{x}+tw')\|_{\ell_1/\ell_2} - \|D^*\bar{x}\|_{\ell_1/\ell_2}(\bar{x}) - t\langle D\bar{z}, w'\rangle}{\frac{1}{2}t^2} \\ &= \liminf_{t\downarrow 0,\, w'\to w} \left(\frac{\|D_T^*(x_0+tw')\|_{\ell_1/\ell_2} - \|D_T^* x_0\|_{\ell_1/\ell_2} - t\langle De, w'\rangle}{\frac{1}{2}t^2} + \frac{-\langle D_S\bar{z}, w'\rangle + \|D_S^* w'\|_{\ell_1/\ell_2}}{\frac{1}{2}t} \right). \end{aligned}$$

Note further that

$$-\langle D_S\bar{z}, w'\rangle + \|D_S^* w'\|_{\ell_1/\ell_2} = -\langle P_S\bar{z}, D_S^* w'\rangle + \|D_S^* w'\|_{\ell_1/\ell_2} \ge 0$$

as $\|P_S\bar{z}\|_{\ell_\infty/\ell_2} \le 1$. It is similar to (5.16) and (5.17) that

$$\operatorname{dom} d^2 J(\bar{x}|\, D\bar{z}) = \{w \in \mathbb{R}^n|\ -\langle DP_S\bar{z}, w\rangle + \|D_S^* w\|_{\ell_1/\ell_2} = 0\} \stackrel{\text{def}}{=} \mathcal{F}, \tag{5.36}$$

$$d^2 J(\bar{x}|\, D\bar{z})(w) = \sum_{g\in I} \frac{\|(D_T^* w)_g\|^2}{\|(D_T^* x_0)_g\|} - \frac{\langle (D_T^* w)_g, (D_T^* x_0)_g\rangle^2}{\|(D_T^* x_0)_g\|^3}. \tag{5.37}$$

Since $\nabla^2 f(\Phi\bar{x}) \succ 0$, we derive from (5.35), (5.36), and (5.37) that

$$\operatorname{Ker} d^2\Theta(\bar{x}|0) = \operatorname{Ker}\Phi \cap \mathcal{E} \cap \mathcal{F}.$$

Furthermore, with $w \in \operatorname{Ker}\Phi$ we have $\Phi w = 0$ and

$$\langle DP_S\bar{z}, w\rangle = \langle D\bar{z}, w\rangle - \langle De, w\rangle = \langle -\Phi^*\nabla f(\Phi\bar{x}), w\rangle - \langle De, w\rangle = -\langle De, w\rangle.$$

This together with (5.36) implies that

$$\operatorname{Ker} d^2\Theta(\bar{x}|0) = \operatorname{Ker}\Phi \cap \mathcal{E} \cap \mathcal{F} = \operatorname{Ker}\Phi \cap \mathcal{E} \cap \operatorname{bd}\mathcal{C}_J(x_0) = \{0\}.$$

By Lemma 2.4, $\bar{x}$ is a strong solution of problem (5.34). The proof is complete. □

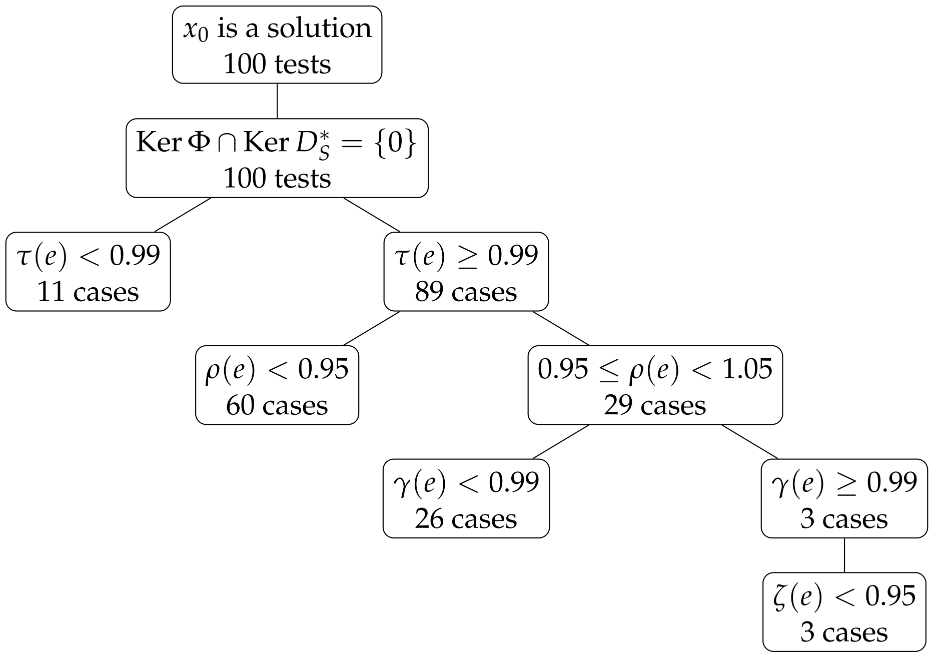
6 Numerical verification of solution uniqueness for group-sparsity

With an aim of demonstrating that our conditions for sharp, strong/unique solution are verifiable, we have implemented a simulation using synthetic data. In our simulation study, Φ was generated as an (2000×240) Gaussian matrix whose entries are independently and identically drawn from the standard normal distribution, $\mathcal{N}(0,1)$. We next randomly divided the set of indicators range from 1 to 2000 into $k = 100$ groups of size 20 with 5 randomly selected active groups. Then, we constructed a measured signal of length $m = 240$, $y_0 \stackrel{\text{def}}{=} \Phi x_0$, based on the original signal x_0 whose elements in each active group are independently and identically distributed $\mathcal{N}(0,1)$. We then used our proposed conditions to verify whether x_0 is a solution of (1.2) using the conditions in Proposition 4.4 and summarize the number of cases where x_0 is classified as sharp or unique/strong solution by the criteria from Theorem 4.6 and 5.3. For checking strong and sharp minima, we only need to compute the Source Coefficient $\rho(e)$ and the Strong Source Coefficient $\zeta(e)$ whenever $\tau(e)$ in (4.21) or $\gamma(e)$ in (5.30) is greater than or equal to 1, respectively, since calculating these numbers are much easier. Similar to the scheme for calculating $\zeta(e)$ in (5.28), $\rho^2(e)$ is the optimal value to the following convex optimization problem (recall (4.11))

$$\min_{t\geq 0, z} t \quad \text{such that} \quad NDz = -NDe, \quad \|z_g\|^2 \leq t, g \in K, \ \text{and} \ z \in \bigoplus_{g\in K} V_g, \tag{6.1}$$

where K is the set of inactive groups. Note that (5.28) and (6.1) are second-order cone programming problems and can be solved via function `solvers.socp` of `cvxopt` package. In our experiment, $\rho(e)$ or $\zeta(e)$ are calculated when $\tau(e)$ or $\gamma(e)$ is greater than or equal to 0.99.

The results are recorded in the following tree diagram.



In all 100 tested random cases, x_0 is verified as a solution of (1.2) and satisfying the Restricted Injectivity Condition. Among them, there are 11 cases with $\tau(e) < 0.99$ thus are classified as sharp solution, the rest are passed to next step for calculating $\rho(e)$. There are 60 cases with $\rho < 0.95$ and 29 tests with $0.95 < \rho(e) < 1.05$. Hence, we had 71 cases x_0 is the sharp solution. We continue the experiment by checking the strong solution condition on the rest 29 cases. Note that since all cases satisfy the Restricted Injectivity Condition, they automatically satisfy the Strong Restricted Injectivity Condition. It remains to check the Analysis Nondegenerate Source Condition. All 29 cases are indicated as satisfying the Analysis Nondegenerate Source Condition with 26 cases having $\gamma(e) < 0.99$ and 3 cases having $\gamma(e) \geq 0.99$ and $\zeta(e) < 0.95$. It means that we have 29 cases of strong solutions that are not sharp.

	number of cases
Sharp solution	71
Strong solution (non-sharp)	29

Table 1: Number of cases with strong and sharp solutions

7 Conclusion

In this paper we show that sharp minima and strong minima play important roles for robust recovery with different rates. We also provide some quantitative characterizations for sharp solutions to convex regularized problems. Unique solutions to ℓ_1 problems are actually sharp solutions. For group sparsity problems, unique solutions are strong solutions. We also obtain several conditions guaranteeing solution uniqueness to group sparsity problems. As solution uniqueness to ℓ_1 problem plays a central role in the area of exact recovery with high probability, we plan to use our results to find a better bound for exact recovery for group-sparsity problems in comparison with the one obtained in [8, 32], at which they only use sufficient conditions for solution uniqueness.

Example 4.15 and Example 5.11 raise important questions about the solution uniqueness and strong minima for the nuclear norm minimization problems. Unique solutions to (1.2) with the nulear norm are neither sharp nor strong solutions. But a strong solution in this case is certainly a unique solution. It means that second-order analysis can provide a sufficient condition for solution uniqueness, and such a condition should be weaker than the one in Theorem 4.6 for sharp solutions. However, unlike the analysis in Lemma 5.2, the second subderivative of the nuclear norm is far more intricate to compute. Understanding solution uniqueness and strong minima for the case of the nuclear norm, or more generally for spectral functions, is a project that we plan to pursue in the future.

Acknowledgements. The authors are indebted to both anonymous referees for their careful readings and thoughtful suggestions that allowed us to improve the original presentation significantly.

Data Availability Statement. *The data underlying this article will be shared on reasonable request to the corresponding author*.

References

[1] D. Amelunxen, M. Lotz, M. B. McCoy, and J. A. Tropp. Living on the edge: phase transitions in convex programs with random data. *Inf. Inference*, 3(3):224–294, 2014.

[2] A. Auslender and M. Teboulle. *Asymptotic cones and functions in optimization and variational inequalities*. Springer Monographs in Mathematics. Springer-Verlag, New York, 2003.

[3] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011. With a foreword by Hédy Attouch.

[4] Y. Bello-Cruz, G. Li, and T. T. A. Nghia. On the linear convergence of forward-backward splitting method: Part I—Convergence analysis. *J. Optim. Theory Appl.*, 188(2):378–401, 2021.

[5] J. F. Bonnans and A. Shapiro. *Perturbation analysis of optimization problems*. Springer Series in Operations Research. Springer-Verlag, New York, 2000.

[6] A. M. Bruckstein, D. L. Donoho, and M. Elad. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.*, 51(1):34–81, 2009.

[7] A. Blake and A. Zisserman. *Visual reconstruction*. MIT press, 1987.

[8] E. Candès and B. Recht. Simple bounds for recovering low-complexity models. *Math. Program.*, 141(1-2, Ser. A):577–589, 2013.

[9] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6):717–772, 2009.

[10] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.

[11] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky. The convex geometry of linear inverse problems. *Found. Comput. Math.*, 12(6):805–849, 2012.

[12] L. Cromme. Strong uniqueness. A far-reaching criterion for the convergence analysis of iterative procedures. *Numer. Math.*, 29(2):179–193, 1977/78.

[13] D. L. Donoho and X. Huo. Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Inform. Theory*, 47(7):2845–2862, 2001.

[14] C. Dossal and R. Tesson. Consistency of ℓ_1 recovery from noisy deterministic measurements. *Applied and Computational Harmonic Analysis*, 36(3):508–513, 2014.

[15] J. Fadili, G. Peyré, S. Vaiter, C.-A. Deledalle, and J. Salmon. Stable recovery with analysis decomposable priors. In *In International Conference on Sampling Theory and Applications (SampTA)*, Bremen, 2013.

[16] S. Foucart and H. Rauhut. *A mathematical introduction to compressive sensing*. Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, New York, 2013.

[17] J. J. Fuchs. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Trans. Inform. Theory*, 51(10):3601–3608, 2005.

[18] J. C. Gilbert. On the solution uniqueness characterization in the L1 norm and polyhedral gauge recovery. *J. Optim. Theory Appl.*, 172(1):70–101, 2017.

[19] M. Grasmair. Linear convergence rates for Tikhonov regularization with positively homogeneous functionals. *Inverse Problems*, 27(7):075014, 16, 2011.

[20] M. Grasmair, M. Haltmeier, and O. Scherzer. Necessary and sufficient conditions for linear convergence of ℓ^1-regularization. *Comm. Pure Appl. Math.*, 64(2):161–182, 2011.

[21] B. Hofmann, B. Kaltenbacher, C. Pöschl, and O. Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.

[22] M. Hebiri and S. Van De Geer. The smooth-Lasso and other $\ell_1 + \ell_2$ penalized methods. *Electronic Journal of Statistics*, 5, 1184–1226, 2011.

[23] W. Li. Error bounds for piecewise convex quadratic programs and applications. *SIAM Journal on Control and Optimization*, 33(5), 1510–1529, 1995.

[24] J. S. Jørgensen, C. Kruschel, and D. A. Lorenz. Testable uniqueness conditions for empirical assessment of undersampling levels in total variation-regularized X-ray CT. *Inverse Probl. Sci. Eng.*, 23(8):1283–1305, 2015.

[25] C. Meyer. *Matrix analysis and applied linear algebra*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. With 1 CD-ROM (Windows, Macintosh and UNIX) and a solutions manual (iv+171 pp.).

[26] B. Mordukhovich and R. Rockafellar. Second-order subdifferential calculus with applications to tilt stability in optimization. *SIAM Journal on Optimization*, 22(3), 953–986, 2011.

[27] S. Mousavi and J. Shen. Solution uniqueness of convex piecewise affine functions based optimization with applications to constrained ℓ_1 minimization. *ESAIM Control Optim. Calc. Var.*, 25 (56), 2019.

[28] S. Nam, M. E. Davies, M. Elad, and R. Gribonval. The cosparse analysis model and algorithms. *Appl. Comput. Harmon. Anal.*, 34(1):30–56, 2013.

[29] S. N. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu. A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. *Statist. Sci.*, 27(4):538–557, 2012.

[30] B. T. Polyak. Sharp minima. Technical report, Institute of Control Sciences Lecture Notes, Moscow, USSR, 1979.

[31] B. T. Polyak. *Introduction to optimization*. Translations Series in Mathematics and Engineering. Optimization Software, Inc., Publications Division, New York, 1987. Translated from the Russian, With a foreword by Dimitri P. Bertsekas.

[32] N. Rao, B. Recht, and R. Nowak. Universal measurement bounds for structured sparse signal recovery. In *Proceedings of AISTATS*, 2012.

[33] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[34] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, 1976.

[35] R. T. Rockafellar and R. J.-B. Wets. *Variational analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.

[36] V. Roth and B. Fischer. The group-lasso for generalized linear models: uniqueness of solutions and efficient algorithms. In *Proceedings of the 25th international conference on Machine learning*, 2008.

[37] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen. *Variational methods in imaging*, volume 167. Springer, 2009.

[38] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Roy. Statist. Soc. Ser. B*, 58(1):267–288, 1996.

[39] J. A. Tropp. Convex recovery of a structured signal from independent random linear measurements,. In *Sampling theory, a renaissance*, Appl. Numer. Harmon. Anal., pages 67–102. Birkhäuser/Springer, Cham, 2015.

[40] J. A. Tropp. Greed is good: algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.

[41] J. A. Tropp. Just relax: convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inform. Theory*, 52(3):1030–1051, 2006.

[42] S. Vaiter. *Low Complexity Regularization of Inverse Problems*. PhD thesis, Université Paris Dauphine - Paris IX, 2014.

[43] S. Vaiter, M. Golbabaee, J. Fadili, and G. Peyré. Model selection with low complexity priors. *Inf. Inference*, 4(3):230–287, 2015.

[44] S. Vaiter, G. Peyré, C. Dossal, and J. Fadili. Robust sparse analysis regularization. *IEEE Trans. Inform. Theory*, 59(4):2001–2016, 2013.

[45] S. Vaiter, G. Peyré, and J. Fadili. Low complexity regularization of linear inverse problems. In *Sampling theory, a renaissance*, Appl. Numer. Harmon. Anal., pages 103–153. Birkhäuser/Springer, Cham, 2015.

[46] M. Yuan and Y. Lin. Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68(1):49–67, 2006.

[47] H. Zou and T. Hastie. Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. Ser. B*, 67 (2), 301–320, 2005.

[48] H. Zhang, M. Yan, and W. Yin. One condition for solution uniqueness and robustness of both ℓ_1-synthesis and ℓ_1-analysis minimizations. *Adv. Comput. Math.*, 42(6):1381–1399, 2016.

[49] H. Zhang, W. Yin, and L. Cheng. Necessary and sufficient conditions of solution uniqueness in 1-norm minimization. *J. Optim. Theory Appl.*, 164(1):109–122, 2015.