LASSO reloaded: a variational analysis perspective with applications to compressed sensing*

Aaron Berk[†], Simone Brugiapaglia[‡], and Tim Hoheisel[§]

Abstract. This paper provides a variational analysis of the unconstrained formulation of the LASSO problem, ubiquitous in statistical learning, signal processing, and inverse problems. In particular, we establish smoothness results for the optimal value as well as Lipschitz properties of the optimal solution as functions of the right-hand side (or *measurement vector*) and the regularization parameter. Moreover, we show how to apply the proposed variational analysis to study the sensitivity of the optimal solution to the tuning parameter in the context of compressed sensing with subgaussian measurements. Our theoretical findings are validated by numerical experiments.

Key words. Variational analysis, LASSO, compressed sensing, convex optimization, Lipschitz stability, subgaussian matrix

AMS subject classifications. 49J53, 62J07, 90C25, 94A12, 94A20

1. Introduction.

One of the most important problems in the applied mathematical sciences is to recover a signal $x \in \mathbb{R}^n$ from noisy linear measurements $b = Ax + e \in \mathbb{R}^n$, where $A \in \mathbb{R}^{m\times n}$ is a measurement (or sensing) matrix and $e \in \mathbb{R}^m$ is a noise vector. A fundamental observation is that such a *linear inverse problem* can be assumed (or cast to) have *sparse solutions*, which can be recovered with high probability from $m \ll n$ (random) observations via computationally efficient signal reconstruction strategies. This is well documented in the groundbreaking work by Donoho [14] and Candès, Romberg, and Tao [11, 12], which gave rise to the field of *compressed sensing*. Since its introduction, the compressed sensing paradigm led to major technological advances in a vast array of signal processing applications, such as, most notably, compressive imaging. For an introduction to the field, its applications, and historical remarks, we refer the reader to [2, 16, 18, 28, 45].

In the noiseless setting (i.e., when $e = 0$), the sparse recovery paradigm for linear inverse problems manifests itself in the optimization framework

$$\min_{z\in\mathbb{R}^n} \|z\|_0 \text{ s.t. } Az = b, \tag{1.1}$$

where $\|\cdot\|_0$ is counting the nonzero entries of a vector in $\mathbb{R}^n$. Despite the absence of noise, problem (1.1) is provably NP-hard in general [18, 33]. Thus, many convex relaxations of

---

*Submitted to the editors DATE.

Funding: The first author was partially supported by a postdoc stipend from the *Centre de Recherche Mathématiques (CRM)* as well as the *Institut de valorisation des données* (IVADO) and NSERC. The second author acknowledges the support of NSERC through grant RGPIN-2020-06766, the Faculty of Arts and Science of Concordia University and the CRM. The third author was partially supported by the NSERC discovery grant RGPIN-2017-04035.

[†]Dept. Math & Stats, McGill University, Montréal, QC, Canada (aaron.berk@mcgill.ca)

[‡]Dept. Math & Stats, Concordia University, Montréal, QC, Canada (simone.brugiapaglia@concordia.ca)

[§]Dept. Math & Stats, McGill University, Montréal, QC, Canada (tim.hoheisel@mcgill.ca)

this optimization problem have been proposed, all of which, in essence, rely on the fact that the l_1-norm is the *convex envelope* of $\|\cdot\|_0$ restricted to an l_∞-ball (see, e.g., [2, §D.4]).

Here we focus on sparse recovery via l_1 minimization in the noisy setting (i.e., when $e \neq 0$) based on the well-known *(unconstrained) LASSO (Least Absolute Shrinkage and Selection Operator)* problem

$$\min_{z \in \mathbb{R}^n} \frac{1}{2}\|Az - b\|^2 + \lambda \|z\|_1, \tag{1.2}$$

where $\lambda > 0$ is a regularization (or tuning) parameter and where $\|\cdot\|$ and $\|\cdot\|_1$ denote the l_2- and l_1-norm, respectively. The LASSO problem was originally proposed (in a constrained formulation) by Tibshirani in the context of statistical regression in the pioneering paper [42]. Since then, the LASSO has become an indispensable tool in statistical learning, especially when performing tasks such as model selection (see [21] and references therein). Moreover, it plays a key role in Bayesian statistics, thanks to its ability to characterize maximum *a posteriori* estimators in linear regression when the parameters have independent Laplace priors [35]. We also note that problem (1.2) was introduced in signal processing by Chen, Donoho, and Saunders in [13] under the name of *basis pursuit denoising*. Here on, we refer to the unconstrained LASSO simply as the LASSO.

From the optimization perspective, the LASSO falls into the category of (additive) composite problems, for which many numerical solution methods have been devised and which have been tested on (1.2), including proximal gradient methods (e.g., FISTA [8]) and primal-dual methods, see Beck's excellent textbook [5] for references, or proximal Newton-type methods proposed by Lee, *et al.* [29], Khan *et al.* [27], Kanzow and Lechner [26] or Milzarek and Ulbrich [30].

From the perspective of *variational analysis* [10, 15, 31, 32, 37], given any optimization problem with parameters, the question as to the behavior of the optimal value and the optimal solution(s) as functions of the parameters arises naturally. The trifecta for solutions of any optimization problem is: *existence, uniqueness* and *stability*. For the LASSO problem, existence is easily established as the l_1-norm is *coercive* (and the quadratic term is not *counter-coercive*[1]). Sufficient conditions for uniqueness were established by Tibshirani [43] and Fuchs [20, Theorem 1]. A set of conditions (see Assumption 4.1 below) that characterizes uniqueness of solutions of a whole class of l_1-optimization problems (including the LASSO) were established by Zhang, *et al.* [47]. An alternative, shorter proof (even though it is not explicitly stated for the LASSO) of this characterization was given by Gilbert [22] which relies, in essence, on polyhedral convexity. Stability results for the LASSO in the sense of variational analysis are, to the best of our knowledge, missing thus far.

Nevertheless, previous work has examined sensitivity of various formulations of l_1 optimization techniques with respect to the choice of the tuning parameter; and other work has examined their robustness to, e.g., measurement error [1, 19]. Regarding the selection of the tuning parameter, the choice of the optimal parameter has been well-studied. In the case of LASSO, the optimal choice of tuning parameter was analyzed in [9] and [38] in such a way as to yield a notion of stability for all sufficiently large λ. Other work has

[1] In the sense of [37, Definition 3.25].

characterized the recovery error for LASSO in terms of the tuning parameter, but does not discuss notions of sensitivity [4, 34, 40, 41]. An asymptotic result establishing sensitivity of the error when λ is less than the optimal choice has been exhibited in a closely related simplification [6]. Notions of sensitivity of the recovery error with respect to variation of the tuning parameter have been discussed in previous work for other formulations of the LASSO program [7].

Main contributions. The main contributions of this paper are the following:

- We establish (in Proposition 3.1) the smoothness of optimal value functions for general regularized least-squares problems, which encompass the LASSO problem (1.2) as a special case.
- We demonstrate (in Example 4.5) that the conditions established by Zhang, *et al.* [47, Condition 2.1], which characterize uniqueness of solutions for the LASSO problem (1.2), do not generally suffice to obtain a locally single-valued, let alone locally Lipschitz, solution function.
- Under a set of assumptions that have been previously used to establish uniqueness of solutions [20, Theorem 1], we prove (in Theorem 4.9) local Lipschitz continuity of the solution map

$$(b, \lambda) \in \mathbb{R}^n \times \mathbb{R}_{++} \mapsto \operatorname*{argmin}_{z \in \mathbb{R}^n} \left\{ \frac{1}{2}\|Az - b\|^2 + \lambda \|z\|_1 \right\}.$$

  As an intermediate step, under the same set of assumptions, we prove (in Proposition 4.7) *(strong) metric regularity* of the subdifferential operator of the objective function $\varphi := \frac{1}{2}\|A(\cdot) - b\|^2 + \lambda\| \cdot \|_1$. The metric regularity of φ is of independent interest, as it is used in, e.g., [27] to establish convergence of a numerical method for solving the LASSO problem. In Example 4.12, we provide further insights on the assumptions used to prove these results and on the sharpness of the resulting Lipschitz constant bound under perturbations of λ (see Corollary 4.11).
- Finally, motivated by compressed sensing applications, we show how to apply these results to study the sensitivity of LASSO solutions to the tuning parameter λ when A is a subgaussian random matrix and $m \ll n$. This is first addressed in Proposition 5.4, under an additional assumption on the sparsity of the LASSO solution. In Proposition 5.5 we show how to remove this assumption when $b = Ax + e$ for some s-sparse vector $x \in \mathbb{R}^n$ and under a bounded noise model for e. We also validate our theoretical findings with numerical experiments in Subsection 5.2.

Roadmap. The rest paper is organized as follows: In Section 2, we provide the background from variational and convex analysis necessary for our study. Section 3 is devoted to the convex analysis of optimal values for regularized least-squares problems as a function of the regularization parameter and the right-hand side (or measurement vector). In turn, in Section 4, we study the optimal solution(s) of the LASSO problem as a function of the regularization parameter and the right-hand side through the lens of variational analysis. Section 5 brings the findings from the previous section to bear on compressed sensing with subgaussian random measurements. We close with some final remarks in Section 6.

Notation: We write $\mathbb{R}_+$ for the nonnegative real numbers, $\mathbb{R}_{++}$ for the positive real numbers, and $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ for the extended real line. For a scalar $t \in \mathbb{R}$, its *sign* is denoted by $\mathrm{sgn}(t)$. For a vector $z \in \mathbb{R}^n$, the operation is to be applied component-wise, i.e. $\mathrm{sgn}(z) = (\mathrm{sgn}(z_i))_{i=1}^n$. The *support* of the vector $z \in \mathbb{R}^n$ is $\mathrm{supp}(z) = \{i \in \{1, \dots, n\} \mid z_i \neq 0\}$. The set of all linear maps from the Euclidean space $\mathbb{E}$ into another $\mathbb{E}'$ will be denoted by $\mathcal{L}(\mathbb{E}, \mathbb{E}')$. For a matrix $A \in \mathbb{R}^{m \times n}$ and $I \subseteq \{1, \dots, n\}$, we denote by $A_I \in \mathbb{R}^{m \times |I|}$ the matrix composed of the columns of A corresponding to I.

2. Preliminaries.

In what follows, let $(\mathbb{E}, \langle \cdot, \cdot \rangle)$ be a Euclidean, i.e. a finite-dimensional real inner product space. For our purposes, $\mathbb{E}$ will be a product space of the form $\mathbb{E} = \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$, whose inner product is the sum of the standard Euclidean inner products of the respective factors. We equip $\mathbb{E}$ with the Euclidean norm derived from the inner product through $\|z\| := \sqrt{\langle z, z \rangle}$ for all $z \in \mathbb{E}$. The induced operator norm of $L \in \mathcal{L}(\mathbb{E}, \mathbb{E}')$ is also denoted by $\| \cdot \|$ and given by

$$\|L\| := \max_{\|z\| \leqslant 1} \|L(z)\| = \max_{\|z\| = 1} \|L(z)\|.$$

Moreover, we denote the unit l_p-ball of $\mathbb{R}^n$ as $\mathbb{B}_p$, where the dimension n of the ambient space will be clear from the context. We denote the minimum and maximum singular values of a matrix $B \in \mathbb{R}^{m \times p}$ by $\sigma_{\min}(B)$ and $\sigma_{\max}(B)$, respectively. Recall that the operator (or spectral) norm of $B \in \mathbb{R}^{m \times p}$ is its largest singular value $\sigma_{\max}(B)$ [24]. In particular, the following result holds.

Lemma 2.1. *Let $B \in \mathbb{R}^{m \times (n-r)}$ and let $[B\ 0] \in \mathbb{R}^{m \times n}$. Then $\|[B\ 0]\| = \|B\|$.*

Proof. This follows immediately from the fact that the singular values of B and $[B\ 0]$ are the square roots of the eigenvalues of BB^T. ∎

2.1. Tools from variational analysis.

We provide in this section the necessary tools from variational analysis, and we follow here the notational conventions of Rockafellar and Wets [37], but the reader can find the objects defined here also in the excellent books by Mordukhovich [31, 32] or Dontchev and Rockafellar [15].

The *regular normal cone* of $C \subseteq \mathbb{E}$ at $\bar{x}$ is

$$\hat{N}_C(\bar{x}) = \left\{ v \in \mathbb{E} \;\middle|\; \limsup_{z \to \bar{x}} \frac{\langle v, z - \bar{x} \rangle}{\|z - \bar{x}\|} \leqslant 0 \right\}.$$

The *limiting normal cone* of C at $\bar{x}$ is the *outer limit*[2] of the regular normal cone, i.e.

$$N_C(\bar{x}) = \left\{ v \in \mathbb{E} \;\middle|\; \exists \{z_k\} \to \bar{x}, \{v_k \in \hat{N}_C(z^k)\} : v^k \to v \right\}.$$

When C is convex (locally around $\bar{x}$), both the limiting and regular normal cone coincide with the normal cone of convex analysis, i.e.

$$N_C(\bar{x}) = \hat{N}_C(\bar{x}) = \{ v \in \mathbb{E} \mid \langle v, \bar{x} - z \rangle \leqslant 0 \ \forall z \in C \}.$$

[2] In the sense of Painlevé-Kuratowski [37, Chapter 5]

Let $S : \mathbb{E}_1 \rightrightarrows \mathbb{E}_2$ be a *set-valued map*, i.e. S maps points in $\mathbb{E}_1$ to sets in $\mathbb{E}_2$. The *graph* of S is the set $\operatorname{gph} S := \{(z,u) \in \mathbb{E}_1 \times \mathbb{E}_2 \mid u \in S(z)\}$, while the *domain* of S is the projection of its graph onto the first component, i.e. $\operatorname{dom} S := \{z \in \mathbb{E}_1 \mid S(z) \neq \emptyset\}$. The *inverse* of S is $S^{-1} : \mathbb{E}_2 \rightrightarrows \mathbb{E}_1$ given by $S^{-1}(y) = \{z \in \mathbb{E}_1 \mid y \in S(z)\}$. The *coderivative* of S at $(\bar{x}, \bar{y}) \in \operatorname{gph} S$ is the (set-valued) map $D^*S(\bar{x} \mid \bar{y}) : \mathbb{E}_2 \rightrightarrows \mathbb{E}_1$ defined via

$$v \in D^*S(\bar{x} \mid \bar{y})(y) \quad \Longleftrightarrow \quad (v, -y) \in N_{\operatorname{gph} S}(\bar{x}, \bar{y}).$$

Given a (positively homogenous) map $H : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, its *outer norm* is given by

$$|H|^+ := \sup_{\|z\| \leqslant 1} \sup_{y \in H(z)} \|y\|.$$

2.2. Tools from convex analysis. Let $f : \mathbb{E} \to \overline{\mathbb{R}}$. Its *domain* is the set $\operatorname{dom} f := \{z \in \mathbb{E} \mid f(z) < +\infty\}$. We call f *proper* if $\operatorname{dom} f \neq \emptyset$ and $f > -\infty$. We call f convex if its *epigraph* $\operatorname{epi} f := \{(z, \alpha) \in \mathbb{E} \mid f(z) \leqslant \alpha\}$ is a convex set, and we call f *closed* (or lower semicontinuous) if $\operatorname{epi} f$ is closed. We define

$$\Gamma_0(\mathbb{E}) := \{f : \mathbb{E} \to \mathbb{R} \cup \{+\infty\} \mid f \text{ closed, proper, convex}\}.$$

The *(Fenchel) conjugate* of f is the function $f^* : \mathbb{E} \to \overline{\mathbb{R}}$ given by

$$f^*(y) = \sup_{z \in \mathbb{E}} \{\langle y,\, z\rangle - f(z)\}.$$

As a pointwise supremum of affine (hence closed, convex) functions it is always closed and convex. As it will occur frequently in our study, we point out that $\|\cdot\|_1^* = \delta_{\mathbb{B}_\infty}$, i.e. the conjugate of the l_1-norm is the indicator function of the l_∞-ball. Here, given a set $S \subset \mathbb{E}$, the *indicator function* of S is defined by

$$\delta_S(z) = \begin{cases} 0, & z \in S, \\ +\infty, & \text{otherwise}, \end{cases} \qquad \forall z \in \mathbb{E}.$$

The *(convex) subdifferential* of f at $\bar{x} \in \mathbb{E}$ is the set given by

$$\partial f(\bar{x}) := \{v \in \mathbb{E} \mid f(\bar{x}) + \langle v,\, z - \bar{x}\rangle \leqslant f(z)\ \forall z \in \mathbb{E}\},$$

which is always closed and convex, possibly empty (even for $\bar{x} \in \operatorname{dom} f$). Note that the subdifferential of the indicator of a convex set $S \subseteq \mathbb{E}$ is the normal cone to S, i.e. $\partial \delta_S = N_S$. The subdifferential operator induces a set-valued map $\partial f : \mathbb{E} \rightrightarrows \mathbb{E}$ which, for $f \in \Gamma_0(\mathbb{E})$, has closed graph, whose (set-valued) inverse is ∂f^*, and whose domain is nonempty and contained in the domain of f, i.e.

$$\emptyset \neq \operatorname{dom} \partial f \subseteq \operatorname{dom} f \quad \forall f \in \Gamma_0(\mathbb{E}).$$

An important example for our study is the l_1-norm $\|\cdot\|_1 : \mathbb{R}^n \to \mathbb{R}$. In this case, we have

$$\partial \|\cdot\|_1(z) = \prod_{i=1}^n \begin{Bmatrix} \operatorname{sgn}(z_i), & z_i \neq 0 \\ [-1,1], & z_i = 0 \end{Bmatrix} = \{y \in \mathbb{B}_\infty \mid \langle z,\, y\rangle = \|z\|_1\}, \quad \forall z \in \mathbb{R}^n. \tag{2.1}$$

The central result that we will use to study optimal value functions of parameterized convex optimization problems is the following.

Theorem 2.2 (Conjugate and subdifferential of optimal value function). *For a function $\psi \in \Gamma_0(\mathbb{E}_1 \times \mathbb{E}_2)$, the optimal value function*

$$p : z \in \mathbb{E}_1 \mapsto \inf_{u \in \mathbb{E}_2} \psi(z, u) \tag{2.2}$$

is convex and the following hold:

(a) *$p^* = \psi^*(\cdot, 0)$, which is closed and convex;*
(b) *for $\bar{x} \in \mathbb{E}_1$ and $\bar{u} \in \operatorname{argmin} \psi(\bar{x}, \cdot)$, we have $\partial p(\bar{x}) = \{v \in \mathbb{E}_1 \mid (v, 0) \in \partial\psi(\bar{x}, \bar{u})\}$;*
(c) *$p^* \in \Gamma_0(\mathbb{E}_1)$ if and only if $\operatorname{dom} \psi^*(\cdot, 0) \neq \emptyset$;*
(d) *$p \in \Gamma_0(\mathbb{E}_1)$ if $\operatorname{dom} \psi^*(\cdot, 0) \neq \emptyset$, and hence the infimum in (2.2) is attained when finite.*

Proof. All statements, except (b), can be found in, e.g., [23, Theorem 3.101]. Part (b) follows from (a) and the fact that, given $(z, y) \in \mathbb{E}_1 \times \mathbb{E}_2$, a pair $(r, s) \in \mathbb{E}_1 \times \mathbb{E}_2$ satisfies $\psi(z, y) + \psi^*(r, s) = \langle (z, y), (r, s) \rangle$ if and only if $(r, s) \in \partial\psi(z, y)$, see, e.g., [36, Theorem 23.5]. ■

3. The value function for regularized least-squares problems.

For $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ and $\lambda > 0$ consider the regularized least-squares problem

$$\min_{z\in\mathbb{R}^n} \frac{1}{2}\|Az - b\|^2 + \lambda R(z), \tag{3.1}$$

where $R \in \Gamma_0(\mathbb{R}^n)$ is a regularizer. The value function for (3.1) is

$$p : (b, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+ \mapsto \inf_{z\in\mathbb{R}^n} \left\{ \frac{1}{2}\|Az - b\|^2 + \lambda R(z) \right\}, \tag{3.2}$$

where we set $0 \cdot R := \delta_{\mathrm{cl}\,(\mathrm{dom}\, R)}$ [3]. The next result studies this value function in depth. Here, note that the *linear image* $A \cdot f$ of a function $f \in \Gamma_0(\mathbb{R}^n)$ under $A \in \mathbb{R}^{m\times n}$ is the convex function given by

$$(A \cdot f)(w) = \inf_{z\in\mathbb{R}^n} \{f(z) \mid Az = w\}, \quad \forall w \in \mathbb{R}^m,$$

which is paired in duality with $f^* \circ A^T$, see e.g. [36, Theorem 16.3]. Moreover, we employ the *recession function* [36] (also called *horizon function* [37]) $f^\infty \in \Gamma_0(\mathbb{R}^n)$ of a convex function $f \in \Gamma_0(\mathbb{R}^n)$ which, given *any* $\bar{x} \in \operatorname{dom} f$, is defined by

$$f^\infty(z) := \sup_{t>0} \frac{f(\bar{x} + tz) - f(\bar{x})}{t}, \quad \forall z \in \mathbb{R}^n.$$

Proposition 3.1 (Regularized least squares). *The following hold for the regularized least-squares problem (3.1):*

(a) *(Existence of solutions) If $R^\infty(z) > 0$ for all $z \in \ker A\backslash\{0\}$, then (3.1) has a solution for all $(b, \lambda) \in \mathbb{R}^m \times \mathbb{R}_{++}$.*

[3] A convention which is backed up by the fact that $\lambda \cdot R$ *epigraphically* converges to $\delta_{\mathrm{cl}\,(\mathrm{dom}\, R)}$ as $\lambda \downarrow 0$. See, e.g., [17, Proposition 4(b)].

(b) *(Differentiability of value function) Let* $(\bar{b}, \bar{\lambda}) \in \mathbb{R}^m \times \mathbb{R}_{++}$ *and let* $\bar{x}$ *be a corresponding solution of* (3.1). *Then the value function* p *in* (3.2) *is differentiable at* $(\bar{b}, \bar{\lambda})$ *with*

$$\nabla p(\bar{b}, \bar{\lambda}) = \begin{pmatrix} \bar{b} - A\bar{x} \\ R(\bar{x}) \end{pmatrix}.$$

Moreover, if (3.1) *has a solution for all* (b, λ) *in some neighborhood* U *of* $(\bar{b}, \bar{\lambda})$, *then* p *is continuously differentiable on* U.

(c) *(Continuity of value function to the boundary) Assume that* $\operatorname{rge} A^T \cap \operatorname{dom} R \neq \emptyset$. *Then* p *is continuous at* $(\bar{b}, 0)$ *for any* $\bar{b} \in \mathbb{R}^m$ *in the sense that*

$$p(b, \lambda) \to \inf_{z \in \operatorname{cl}(\operatorname{dom} R)} \frac{1}{2}\|Az - \bar{b}\|^2 \quad \text{as} \quad (b, \lambda) \to (\bar{b}, 0).$$

(d) *(Convexity of value function)* p *is convex as a function of* b, *concave as a function of* λ.

(e) *(Constancy of residual and regularizer value) Given* $(\bar{b}, \bar{\lambda}) \in \mathbb{R}^m \times \mathbb{R}_{++}$, *the residual* $\bar{b} - A\bar{x}$ *and regularizer value* $R(\bar{x})$ *do not depend on the particular solution* $\bar{x}$.

Proof. (a) Set $\phi := \frac{1}{2}\|A(\cdot) - b\|^2 + \lambda R \in \Gamma_0(\mathbb{R}^n)$. Now observe, cf. [37, p. 89], that $(\frac{1}{2}\|A(\cdot) - b\|^2)^\infty = \delta_{\ker A} + \langle A^T b, \cdot \rangle$. Hence, $\phi^\infty = \delta_{\ker A} + \langle b, A(\cdot)\rangle + \lambda R^\infty$, see [37, Exercise 3.29]. Consequently, using the given assumptions, we have $\phi^\infty(z) > 0$ for all $z \neq 0$, and consequently ϕ is (level-)coercive, see [37, Theorem 3.26], thus admits a minimizer.

(b) For $\lambda > 0$, we observe that $p(b, \lambda) = \lambda v(b, \lambda)$, where

$$v(b, \lambda) = \inf_{z \in \mathbb{R}^n} \left\{ \frac{1}{2\lambda}\|Az - b\|^2 + R(z) \right\}.$$

The latter fits the pattern of [17, Theorem 2] with $\omega := \frac{1}{2}\|\cdot\|^2$, $f := R$ and $L(x, b) = Az - b$. Given any $\bar{x} \in \operatorname{argmin}_{z \in \mathbb{R}^n} \left\{\frac{1}{2}\|Az - b\|^2 + \lambda R(z)\right\}$ it follows from this result[4] that

$$\begin{aligned}\partial v(b, \lambda) &= \left\{ (v, -\frac{1}{2}\|y\|^2) \;\middle|\; y = \frac{1}{\lambda}(A\bar{x} - b),\; -A^T y \in \partial R(\bar{x}), v = -y \right\} \\ &= \left\{ \left(\frac{b - A\bar{x}}{\lambda}, -\frac{1}{2}\left\| \frac{A\bar{x} - b}{\lambda} \right\|^2 \right) \right\}.\end{aligned}$$

Therefore, since v is convex, v is differentiable at (b, λ) with $\nabla v(b, \lambda) = \left(\frac{b - A\bar{x}}{\lambda}, -\frac{1}{2}\left\|\frac{A\bar{x} - b}{\lambda}\right\|^2 \right)^T$, see e.g., [36, Theorem 25.2]. Hence, by the chain rule, we find that p is differentiable at (b, λ) with

$$\nabla p(b, \lambda) = \begin{pmatrix} b - A\bar{x} \\ -\frac{1}{2\lambda}\|A\bar{x} - b\|^2 \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{1}{2\lambda}\|A\bar{x} - b\|^2 + R(\bar{x}) \end{pmatrix} = \begin{pmatrix} b - A\bar{x} \\ R(\bar{x}) \end{pmatrix}.$$

The addendum about continuous differentiability follows readily from [36, Corollary 25.5.1]

[4] Alternatively, this could be derived also from Theorem 2.2.

(c) First note that, by the posed assumptions, we have $A \cdot R \in \Gamma_0(\mathbb{R}^n)$, see [36, Theorem 16.3]. Now, using the *Moreau envelope* [37, Definition 1.22] and *epigraphical multiplication* [37, Exercise 1.28], we observe that

$$\begin{aligned} p(b,\lambda) &= -\inf_{y\in\mathbb{R}^m}\left\{\frac{1}{2}\|y\|^2 - \langle b,\, y\rangle + \lambda\star R^*(A^Ty)\right\} \\ &= \frac{1}{2}\|b\|^2 - \inf_{y\in\mathbb{R}^m}\left\{\lambda\star(R^*\circ A^T)(y) + \frac{1}{2}\|y-b\|^2\right\} \\ &= \frac{1}{2}\|b\|^2 - e_1(\lambda\star(R^*\circ A^T))(b) \\ &= e_1(\lambda(A\cdot R))(b) \\ &= \lambda e_\lambda(A\cdot R)(b) \\ &\to \frac{1}{2}d^2_{\mathrm{cl}\,(A\cdot \mathrm{dom}\,R)}(\bar{b}) \quad \text{as} \quad (b,\lambda)\to(\bar{b},0). \end{aligned}$$

Here the first identity is due to Fenchel-Rockafellar duality [37, Example 11.41], the fourth uses [37, Example 11.26], and the fifth follows from the definition of the Moreau envelope and the fact that $\mathrm{dom}\,(A\cdot R) = A\mathrm{dom}\,R$. The limit property uses [17, Proposition 4(c)]. Realizing that $\frac{1}{2}d^2_{\mathrm{cl}\,(A\cdot\mathrm{dom}\,R)}(\bar{b}) = \inf_{z\in\mathrm{cl}\,(\mathrm{dom}\,R)}\frac{1}{2}\|Az-\bar{b}\|^2$ gives the desired statement.

(d) In the proof of part (a) we saw that, for $\lambda > 0$, we have $p(b,\lambda) = \lambda v(b,\lambda)$ where v is (jointly) convex, see e.g. Theorem 2.2. Then the convexity of $p(\cdot,\bar{\lambda})$ $(\bar{\lambda} > 0)$. On the other hand, for $\bar{b}\in\mathbb{R}^m$, $\lambda,\mu > 0$ and $t\in(0,1)$, we have

$$\begin{aligned} &p(\bar{b}, t\lambda + (1-t)\mu) \\ &= \inf_{z\in\mathbb{R}^n}\left\{\frac{1}{2}\|Az-\bar{b}\|^2 + (t\lambda+(1-t)\mu)R(z)\right\} \\ &\geqslant t\cdot\inf_{z\in\mathbb{R}^n}\left\{\frac{1}{2}\|Az-\bar{b}\|^2 + \lambda R(z)\right\} + (1-t)\cdot\inf_{z\in\mathbb{R}^n}\left\{\frac{1}{2}\|Az-\bar{b}\|^2 + \mu R(z)\right\} \\ &= tp(\bar{b},\lambda) + (1-t)p(\bar{b},\mu). \end{aligned}$$

(e) This follows immediately from (b). ■

Remark 3.2. Proposition 3.1 remains valid (with the appropriate adjustments) when $\frac{1}{2}\|\cdot\|^2$ is replaced by a (squared) weighted Euclidean norm $\frac{1}{2}\langle V\cdot,\,\cdot\rangle$ for some symmetric positive definite matrix V. ◇

Using $R = \|\cdot\|_1$ in (3.1) we can state the following immediate result for the LASSO problem.

Corollary 3.3 (LASSO). *The LASSO problem* (1.2) *always has a solution, and for its value function*

$$p : (b,\lambda)\in\mathbb{R}^n\times\mathbb{R}_+ \mapsto \inf_{z\in\mathbb{R}^n}\left\{\frac{1}{2}\|Az-b\|^2 + \lambda\|z\|_1\right\}$$

the following hold:

(a) *Let* $(\bar{b}, \bar{\lambda}) \in \mathbb{R}^m \times \mathbb{R}_{++}$ *and let* $\bar{x}$ *be a corresponding solution of* (1.2). *Then* p *is continuously differentiable at* $(\bar{b}, \bar{\lambda})$ *with*

$$\nabla p(\bar{b}, \bar{\lambda}) = \begin{pmatrix} \bar{b} - A\bar{x} \\ \|\bar{x}\|_1 \end{pmatrix}.$$

In particular, given $(\bar{b}, \bar{\lambda}) \in \mathbb{R}^m \times \mathbb{R}_{++}$, *the residual* $\bar{b} - A\bar{x}$ *and regularizer value* $\|\bar{x}\|_1$ *do not depend on the particular solution* $\bar{x}$.

(b) *For any* $\bar{b} \in \mathbb{R}^m$, $p(b, \lambda) \to \frac{1}{2} d^2_{\mathrm{rge}\, A}(\bar{b})$ *as* $(b, \lambda) \to (\bar{b}, 0)$.

As another special case of (3.1), we consider the case $R = \frac{1}{2}\|\cdot\|^2$ which is known as Tikhonov regularization.

Corollary 3.4 (Tikhonov regularization). *Consider*

$$\min_{z \in \mathbb{R}^n} \frac{1}{2}\|Az - b\|^2 + \frac{\lambda}{2}\|z\|^2. \tag{3.3}$$

Then the following hold for the value function

$$p : (b, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+ \mapsto \inf_{z \in \mathbb{R}^n} \left\{ \frac{1}{2}\|Az - b\|^2 + \frac{\lambda}{2}\|z\|^2 \right\}.$$

(a) p *is continuously differentiable on* $\mathbb{R}^m \times \mathbb{R}_{++}$ *with*

$$\nabla p(\bar{b}, \bar{\lambda}) = \begin{pmatrix} \bar{b} - Ax(b, \lambda) \\ \frac{1}{2}\|x(b, \lambda)\|^2 \end{pmatrix},$$

where $x(b, \lambda) = (\lambda A^T A + I)^{-1}(A^T b)$.

(b) *For any* $\bar{b} \in \mathbb{R}^m$, $p(b, \lambda) \to \frac{1}{2} d^2_{\mathrm{rge}\, A}(\bar{b})$ *as* $(b, \lambda) \to (\bar{b}, 0)$.

The fact that the solution map in the Tikhonov setting can be written out explicitly as $x(b, \lambda) = (\lambda A^T A + I)^{-1}(A^T b)$ is due to the fact that the subdifferential of the regularizer $R = \frac{1}{2}\| \cdot \|^2$ is simply the identity, which is perfectly aligned with the quadratic fidelity term. There is no explicit inversion for general $R \in \Gamma_0(\mathbb{R}^n)$. This provides a nice segue to the following section, where we study the solution map for the l_1-regularizer through implicit function theory provided by variational analysis.

4. The solution map of LASSO.

This section is devoted to the study of the optimal solution function of the LASSO problem (1.2). We start by recalling that, thanks to the analysis by Zhang, *et al.* in [47], a solution $\bar{x}$ of the LASSO problem (1.2) (given A, b, λ) is unique if (and only if) the following set of conditions hold:

Assumption 4.1 ([47, Condition 2.1]). For a minimizer $\bar{x}$ of (1.2) and $I := \mathrm{supp}(\bar{x})$ we have that

(i) A_I has full column rank $|I|$;
(ii) there exists $y \in \mathbb{R}^m$ such that $\|A_{I^C}^T y\|_\infty < 1$ and $A_I^T y = \mathrm{sgn}(\bar{x}_I)$.

Remark 4.2. The convex-analytically inclined reader may find it illuminating to realize that Assumption 4.1 is equivalent to the following set of conditions (see, e.g., Gilbert's paper [22] for an explicit proof):

(i) $\operatorname{rge} A^T + \operatorname{par}(\partial\|\cdot\|_1(\bar{x})) = \mathbb{R}^n$;
(ii) $\operatorname{ri}(\partial\|\cdot\|_1(\bar{x})) \cap \operatorname{rge} A^T \neq \varnothing$.

Here $\operatorname{par}(\partial\|\cdot\|_1(\bar{x}))$ and $\operatorname{ri}(\partial\|\cdot\|_1(\bar{x}))$ are the subspace parallel to and the relative interior of the subdifferential $\partial\|\cdot\|_1(\bar{x})$, respectively. ◇

Building on an example by Zhang, *et al.* [47, p. 113], we will show that the solution uniqueness guaranteed by Assumption 4.1 is not stable to perturbations in the tuning parameter λ. A stronger set of conditions, which has already occurred in the literature (see [20]) as a sufficient condition for uniqueness, is the following:

Assumption 4.3. For a minimizer $\bar{x}$ of (1.2) and $I = I(\bar{x}) := \operatorname{supp}(\bar{x})$ we have
(i) A_I has full column rank $|I|$;
(ii) $\|A_{I^C}^T(b - A_I\bar{x}_I)\|_\infty < \lambda$.

The relation between Assumption 4.3 and Assumption 4.1 is clarified now.

Lemma 4.4. *Assumption 4.3 implies Assumption 4.1*

Proof. First note that, since $A\bar{x} = A_I\bar{x}_I$, and since $\bar{x}$ solves (1.2), $\bar{x}$ satisfies the optimality conditions $A^T(b - A_I\bar{x}_I) \in \lambda\partial\|\cdot\|_1(\bar{x})$. Hence, in view of (2.1), we find $A_I^T(b - A_I\bar{x}_I) = \lambda\operatorname{sgn}(\bar{x}_I)$. Setting $\bar{y} := \frac{b - A_I\bar{x}_I}{\lambda}$ then yields Assumption 4.1 (ii).

We now present the example announced above, which expands on an example in Zhang, *et al.* [47].

Example 4.5. Consider the LASSO problem (1.2) with

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 2 & -2 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

The unique solution for $\bar{\lambda} = 1$ (see [47, p. 113]) is $\bar{x} = (0,\ 1/4,\ 0)^T$ with $I(\bar{x}) = \{2\}$. Indeed, we observe that

$$\bar{x}_I = 1/4, \quad A_I = \begin{pmatrix} 0 \\ 2 \end{pmatrix} \quad \text{and} \quad A_{I^C}^T = \begin{pmatrix} 1 & 0 \\ 2 & -2 \end{pmatrix}.$$

In particular, A_I has full column rank, and setting $\bar{y} := [1/2, 1/2]^T$, we find

$$A_I^T\bar{y} = 1 = \operatorname{sgn}(\bar{x}_I) \quad \text{and} \quad \|A_{I^C}^T\bar{y}\|_\infty = \left\|\begin{pmatrix} 1/2 \\ 0 \end{pmatrix}\right\|_\infty = 1/2 < 1.$$

Therefore, the solution $\bar{x}$ satisfies Assumption 4.1, which confirms its uniqueness. On the other hand, we find that $\|A_{I^C}^T(b - A_I\bar{x}_I)\|_\infty = \left\|\binom{1}{1}\right\|_\infty = 1 = \bar{\lambda}$. Hence, Assumption 4.3 is violated and, as we shall see, uniqueness of the solution is not preserved under small perturbations of λ. Indeed, for $\lambda \in (0,1)$, consider the point $\bar{x}^\lambda := (1-\lambda,\ \frac{2-\lambda}{4},\ 0)^T$ and note that $\bar{x}^\lambda \to \bar{x}$ as $\lambda \to \bar{\lambda}$. Then

$$A^Tb - A^TA\bar{x}^\lambda = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} - \begin{pmatrix} 1-\lambda \\ 2-\lambda \\ -\lambda \end{pmatrix} = \begin{pmatrix} \lambda \\ \lambda \\ \lambda \end{pmatrix} \in \lambda\partial\|\cdot\|_1(\bar{x}^\lambda),$$

and hence $\bar{x}^\lambda$ solves the LASSO problem for any $\lambda \in (0,1)$. Moreover, we have $I(\bar{x}^\lambda) = \{1,2\}$ and, hence, $\operatorname{sgn}(\bar{x}^\lambda_{I(\bar{x}^\lambda)}) = (1,\ 1)^T$. Now, $A^T_{I(\bar{x}^\lambda)}y = \operatorname{sgn}(\bar{x}^\lambda_{I(\bar{x}^\lambda)})$ only admits $y = (1,\ 1/2)^T$ as a solution and $\|A^T_{I(\bar{x}^\lambda)^C}y\|_\infty = 1$. This shows that Assumption 4.1 is violated and, consequently, $\bar{x}^\lambda$ is not the unique solution to (1.2). Indeed, it can be seen that, for any $\lambda \in (0,1)$, the points

$$\begin{pmatrix} 1-\lambda-2t \\ \frac{2-\lambda+4t}{4} \\ t \end{pmatrix}, \quad \forall t \in \left[0, \frac{1-\lambda}{2}\right),$$

solve the LASSO problem (1.2). ◇

A necessary ingredient for our study is the normal cone to the graph of the subdifferential of the l_1-norm.

Lemma 4.6 (Normal cone of $\operatorname{gph} \partial\|\cdot\|_1$). *For $(\bar{x},\bar{u}) \in \operatorname{gph} \partial\|\cdot\|_1$ we have*

$$N_{\operatorname{gph} \partial\|\cdot\|_1}(\bar{x},\bar{u}) = \prod_{i=1}^n \begin{cases} \{0\}\times\mathbb{R}, & \bar{x}_i \neq 0, \bar{u}_i = \operatorname{sgn}(\bar{x}_i), \\ \mathbb{R}_+\times\mathbb{R}_- \cup \{0\}\times\mathbb{R} \cup \mathbb{R}\times\{0\}, & \bar{x}_i = 0, \bar{u}_i = -1, \\ \mathbb{R}_-\times\mathbb{R}_+ \cup \{0\}\times\mathbb{R} \cup \mathbb{R}\times\{0\}, & \bar{x}_i = 0, \bar{u}_i = 1, \\ \mathbb{R}\times\{0\}, & \bar{x}_i = 0, |\bar{u}_i| < 1. \end{cases}$$

Proof. Use [37, Proposition 6.41] and the separability $\partial\|\cdot\|_1(z) = \prod_{i=1}^n \partial|\cdot|(z_i)$. ■

The next result is key to our study, and it establishes (strong) metric regularity of the subdifferential operator of the objective function for LASSO to which we want to imply a (set-valued) implicit function theorem à la Dontchev and Rockafellar [15]. It is also of interest beyond the scope of our study: e.g., the numerical method developed by Khan, *et al.* [27] relies on such property.

Here a set-valued map $S : \mathbb{E}_1 \rightrightarrows \mathbb{E}_2$ is called *metrically regular* at $\bar{x} \in \operatorname{dom} S$ for $\bar{u} \in S(\bar{x})$ if there exist neighborhoods V of $\bar{x}$ and W of $\bar{u}$, respectively, and $\kappa > 0$ such that

$$d_{S^{-1}(u)}(z) \leqslant \kappa \cdot d_{S(z)}(u) \quad \forall z \in V,\ u \in W.$$

In case that S has closed graph (locally around $(\bar{x},\bar{y})$), which is the case for the the subdifferential operator of a closed, proper, convex function, metric regularity can be characterized by finiteness of the outer norm $|D^*S(\bar{x}|\bar{y})^{-1}|^+$ [15, Theorem 4C.2].

If S is a monotone operator, as is the case for us, metric regularity (at $(\bar{x},\bar{u})$) is equivalent to strong metric regularity [15] (see also the interesting paper by Aragón, Artacho and Geoffrey [3]), which means that the inverse map S^{-1} is locally Lipschitz around $\bar{u}$. This, of course, is exactly what we need.

Proposition 4.7. *Let $\bar{x}$ be a solution of* (1.2) *with $I := \operatorname{supp}(\bar{x})$ and such that Assumption 4.3 is satisfied. Then for*

$$T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n, \quad T(z) = \frac{1}{\lambda}A^TAz + (\partial\|\cdot\|_1)(z)$$

the following hold:

(a) $|D^*T\left(\bar{x}|\frac{1}{\lambda}A^Tb\right)^{-1}|^+ = \frac{\lambda}{\sigma_{\min}(A_I)^2}$.
(b) T *is strongly metrically regular at* $\bar{x}$ *for* $\frac{1}{\lambda}A^Tb$.

Proof. (a) Set $\bar{u} := \frac{1}{\lambda}A^T(b - A\bar{x})$ and $H := D^*T\left(\bar{x}|\frac{1}{\lambda}A^Tb\right)^{-1} = \left(\frac{1}{\lambda}A^TA + D^*(\partial\|\cdot\|_1)(\bar{x}|\bar{u})\right)^{-1}$. Then

$$\begin{aligned} y \in H(z) &\iff z - \frac{1}{\lambda}A^TAy \in D^*(\partial\|\cdot\|_1)(\bar{x}|\bar{u})(y) \\ &\iff \begin{cases} \forall i \in I: \left(z_i - \frac{1}{\lambda}a_i^TAy, -y_i\right) \in \{0\} \times \mathbb{R}, \\ \forall i \in I^C: \left(z_i - \frac{1}{\lambda}a_i^TAy, -y_i\right) \in \mathbb{R} \times \{0\}, \end{cases} \\ &\iff y_{I^C} = 0,\ z_I = \frac{1}{\lambda}A_I^TA_Iy_I. \end{aligned}$$

Here the second equivalence uses Lemma 4.6 and Assumption 4.3 (ii) (which implies here that $|\bar{u}_i| < \lambda$ for all $i \in I^C$). We then find that

$$\begin{aligned} |H|^+ &= \sup_{(z,y)\in\mathbb{R}^n\times\mathbb{R}^{|I|}} \left\{ \|y\| \;\middle|\; \|z\| \leqslant 1, z_I = \frac{1}{\lambda}A_I^TA_Iy \right\} \\ &= \lambda \sup_{\|z\|\leqslant 1} \|(A_I^TA_I)^{-1}z_I\| \\ &= \lambda \cdot \|(A_I^TA_I)^{-1}\| \\ &= \lambda \cdot \sigma_{\max}((A_I^TA_I)^{-1}) \\ &= \frac{\lambda}{\sigma_{\min}(A_I)^2}. \end{aligned}$$

Here the second identity uses Assumption 4.3 (i) (i.e. $A_I^TA_I$ is invertible), while the third is due to Lemma 2.1.

(b) From (a) and [15, Theorem 4C.2] it follows that T is metrically regular at $\bar{x}$ for A^Tb. In addition, as a subdifferential of a (closed, proper) convex function, T is (globally, hence locally) monotone, thus by [15, Theorem 3G.5] it follows that T is strongly metrically regular at $\bar{x}$ for A^Tb ■

Before presenting the main result of this section we need an auxiliary statement (whose proof is deferred to the supplement). See the comment below for a slightly weaker result which is readily proven.

Lemma 4.8. *Let* $d \in \mathbb{R}^m$. *Then the operator norm of the matrix* $[I\ d] \in \mathbb{R}^{m\times(m+1)}$ *is given by*

$$\|[I\ d]\| = \sqrt{1 + \|d\|^2}.$$

Let us point out that, in the setting of Lemma 4.8, the slightly weaker estimate $\|[I\ d]\| \leqslant 1 + \|d\|$ follows immediately from the definition of operator norm and the triangle inequality.

We are now in a position to prove our main result. In its proof we utilize the *strict graphical derivative* D_*S of a set-valued map S. Since we only use it here, we refer the reader to the literature [15, 37] for its definition.

Theorem 4.9. *For* $(\bar b, \bar\lambda) \in \mathbb{R}^n \times \mathbb{R}_{++}$ *let* $\bar x$ *be a solution of* (1.2) *with* $I := \operatorname{supp}(\bar x)$. *Under Assumption 4.3, the solution map*

$$S : (b, \lambda) \mapsto \operatorname*{argmin}_{z\in\mathbb{R}^n} \left\{ \frac{1}{2}\|Az - b\|^2 + \lambda\|z\|_1 \right\}$$

is locally Lipschitz at $(\bar b, \bar\lambda)$ *with (local) Lipschitz modulus* L *satisfying*

$$L \leqslant \frac{\sigma_{\max}(A)}{\sigma_{\min}(A_I)^2} \cdot \sqrt{1 + \left\| \frac{A\bar x - \bar b}{\bar\lambda} \right\|^2}.$$

Proof. Consider the function $f : \mathbb{R}^m \times \mathbb{R}_{++} \times \mathbb{R}^n \to \mathbb{R}^n,\ f(b, \lambda, z) = \frac{1}{\lambda} A^T(Az - b)$ and the set-valued map $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n,\ F = \partial\|\cdot\|_1$, and observe that

$$S(b, \lambda) = \{z \in \mathbb{R}^n \mid 0 \in f(b, \lambda, z) + F(z)\} \quad \forall b \in \mathbb{R}^m,\ \lambda > 0.$$

Combining Proposition 4.7 and Exercise 4C.5 in Dontchev and Rockafellar [15], we find that S has the Aubin property at $(\bar b, \bar\lambda)$ for $\bar x$ with

$$\begin{aligned} L &= \left|\left(D_x f(\bar b, \bar\lambda, \bar x) + D^*F(\bar x \mid -f(\bar b, \bar\lambda, \bar x))\right)^{-1}\right|^+ \cdot \limsup_{(b,\lambda,z)\to(\bar b,\bar\lambda,\bar x)} \|D_{(b,\lambda)} f(b, \lambda, z)\| \\ &= \frac{\bar\lambda}{\sigma_{\min}(A_I)^2} \cdot \|D_{(b,\lambda)} f(\bar b, \bar\lambda, \bar x)\| \\ &= \frac{1}{\sigma_{\min}(A_I)^2} \cdot \left\| \left[A^T \quad A^T\!\left(\frac{A\bar x - \bar b}{\bar\lambda}\right) \right] \right\|. \end{aligned}$$

Now, by the compatibility of vector and operator norm and Lemma 4.8 (with $d := \frac{A\bar x - \bar b}{\bar\lambda}$), we obtain

$$\begin{aligned} \left\| \left[A^T \quad A^T\!\left(\frac{A\bar x - \bar b}{\bar\lambda}\right) \right] \right\| &= \max_{\|(v,\alpha)\|=1} \left\| A^T v + \alpha A^T \left(\frac{A\bar x - \bar b}{\bar\lambda}\right) \right\| \\ &\leqslant \max_{\|(v,\alpha)\|=1} \|A^T\| \cdot \left\| v + \alpha \left(\frac{A\bar x - \bar b}{\bar\lambda}\right) \right\| \\ &= \|A\| \cdot \left\| \left[I \quad \frac{A\bar x - \bar b}{\bar\lambda} \right] \right\| \\ &\leqslant \|A\| \cdot \sqrt{1 + \left\| \frac{A\bar x - \bar b}{\bar\lambda} \right\|^2}. \end{aligned}$$

Therefore $L \leqslant \frac{\sigma_{\max}(A)}{\sigma_{\min}(A_I)^2} \cdot \sqrt{1 + \left\| \frac{A\bar x - \bar b}{\bar\lambda} \right\|^2}$.

Now, let $0 \in D_*(f + F)(\bar b, \bar\lambda, \bar x | 0)(0, 0, w)$ or equivalently [37, Exercise 10.43]

$$0 \in \frac{1}{\bar\lambda} A^T A w + D_*(\partial\|\cdot\|_1)(\bar x| - f(\bar b, \bar\lambda, \bar x)) = D_* T\left(\bar x \middle| \frac{1}{\bar\lambda} A^T \bar b\right)(w),$$

where T is the map from Proposition 4.7. Then, by said lemma (part (b)) and the strict derivative criterion for strong metric regularity [15, 37], it follows that $w = 0$. Hence, [37, Theorem 9.56(b)] gives single-valuedness of S, which combined with the Aubin property gives the desired statement. ■

Remark 4.10. When A is considered a parameter, $f(A, b, \lambda, z) = \frac{1}{\lambda}A^T(Az - b)$, one may show (see Remark A.1) the solution map $(A, b, \lambda) \mapsto \operatorname{argmin}_{z\in\mathbb{R}^n}\{\frac{1}{2}\|Az - b\|^2 + \lambda\|z\|_1\}$ is locally Lipschitz at $(\bar{A}, \bar{b}, \bar{\lambda})$ with (local) Lipschitz modulus L satisfying

$$L \leqslant \frac{1}{\sigma_{\min}(\bar{A}_I)^2}\left[\|\bar{A}\bar{x} - \bar{b}\| + \sigma_{\max}\left(\|\bar{x}\| + \sqrt{1 + \left\|\frac{\bar{A}\bar{x} - \bar{b}}{\bar{\lambda}}\right\|^2}\right)\right].$$

◇

When we fix the parameter $\bar{b}$, and look at the solution only as a function of the regularization parameter λ, we can get a siginficantly sharper Lipschitz modulus.

Corollary 4.11. *Under the assumptions of Theorem 4.9 we find that*

$$S : \lambda \mapsto \operatorname*{argmin}_{z\in\mathbb{R}^n}\left\{\frac{1}{2}\|Az - \bar{b}\|^2 + \lambda\|z\|_1\right\} \tag{4.1}$$

is locally Lipschitz at $\bar{\lambda}$ with constant $L < \dfrac{\sqrt{|I| + 1}}{\sigma_{\min}(A_I)^2}$.

Proof. We apply the same reasoning as in the proof of Theorem 4.9 while using that the operator norm of a vector is its Euclidean norm, and observing that

$$\begin{aligned}\|D_\lambda f(\bar{b}, \bar{\lambda}, \bar{x})\| &= \left\|\frac{1}{\bar{\lambda}}\left[\frac{A^T(A\bar{x} - \bar{b})}{\bar{\lambda}}\right]\right\| \\ &= \frac{1}{\bar{\lambda}}\sqrt{\left\|\frac{A_I^T(A_I\bar{x}_I - \bar{b})}{\bar{\lambda}}\right\|^2 + \left\|\frac{A_{I^C}^T(A_I\bar{x}_I - b)}{\bar{\lambda}}\right\|^2} \\ &< \frac{1}{\bar{\lambda}}\sqrt{|I| + 1},\end{aligned}$$

where the last inequality uses Assumption 4.3(ii) and the facts that $A_I^T(A_I\bar{x}_I - \bar{b}) = \bar{\lambda}\operatorname{sgn}(\bar{x}_I)$ and that $\|\operatorname{sgn}(\bar{x}_I)\|^2 = |I|$. ■

4.1. On Assumption 4.3 and the sharpness of Corollary 4.11. We now discuss a case where Assumption 4.3 is satisfied and where the single-valued solution map to the LASSO problem admits an explicit formula. This example is due to Fuchs [20] and is based on the notion of *coherence* (see, e.g. [18, Chapter 5] and references therein). We assume for simplicity that the columns $a_1, \ldots, a_n$ of A have unit l_2-norm and recall that the coherence of A is defined as

$$\mu(A) := \max_{i\neq j}|\langle a_i, a_j\rangle|.$$

Beyond providing more insight on Assumption 4.3, the following example sheds light on the sharpness of the Lipschitz bound proved in Corollary 4.11.

Example 4.12 (Assumption 4.3 and sharpness of Corollary 4.11). Assume that the columns of A have unit l_2-norm and suppose that $\bar{b} \in \mathbb{R}^m$ satisfies

$$\exists x \in \mathbb{R}^n \text{ s.t. } Ax = \bar{b},\ \|x\|_0 < \frac{1}{2}\left(1 + \frac{1}{\mu(A)}\right) \quad \text{and} \quad A_{\text{supp}(z)} \text{ has full rank.} \tag{4.2}$$

In other words, Ax is an irreducible sparse representation of $\bar{b}$ with respect to the columns of A. Then, an inspection of the proofs of [20, Theorem 2] and [20, Theorem 3] reveals that there exists $\lambda_{\max} > 0$ such that Assumption 4.3 is satisfied for all $\lambda \in (0, \lambda_{\max})$. A possible choice of $\lambda_{\max}$ can be made by taking the largest value such that the following condition holds (see [20, Eq. (7)]):

$$\text{sgn}(x_J - \lambda(A_J^T A_J)^{-1}\text{sgn}(x_J)) = \text{sgn}(x_J), \quad \forall \lambda \in [0, \lambda_{\max}),$$

where $J = J(x) := \text{supp}(x)$. In this case, the solution map $(b, \lambda) \mapsto S(b, \lambda)$ of the LASSO problem (1.2) is single valued at $(\bar{b}, \bar{\lambda})$, for every $\bar{\lambda} \in (0, \lambda_{\max})$ and every $\bar{b}$ such that (4.2) holds. In addition, the unique solution $\bar{x} = S(\bar{b}, \bar{\lambda})$ is explicitly characterized by

$$\text{supp}(\bar{x}) = J \quad \text{and} \quad \bar{x}_J = A_J^+ \bar{b} - \bar{\lambda}(A_J^T A_J)^{-1}\text{sgn}(x_J), \tag{4.3}$$

where $A_J^+ = (A_J^T A_J)^{-1} A_J^T$ is the Moore-Penrose pseudoinverse of A_J. We observe that, although $Ax = \bar{b}$, we have $A\bar{x} \neq \bar{b}$. Moreover, the support set J depends on $\bar{b}$ (since it determines which columns of A are used to form the sparse representation of $\bar{b}$). Yet, J is independent of $\bar{\lambda}$. In fact, the existence of x (which, in turn, defines J) in (4.2) is not related to the tuning parameter $\bar{\lambda}$. We note, however, that $\lambda_{\max}$ does depend on J and, hence, on $\bar{b}$. We conclude by observing that the explicit characterization (4.3) yields

$$\|D_\lambda S(\bar{b}, \bar{\lambda})\| = \|(A_J^T A_J)^{-1}\text{sgn}(x_J)\| \leqslant \|(A_J^T A_J)^{-1}\|\|\text{sgn}(x_J)\| = \frac{\sqrt{|J|}}{\sigma_{\min}(A_J)^2}. \tag{4.4}$$

This estimate is consistent with the Lipschitz bound of Corollary 4.11 (up to replacing $\sqrt{|J|}$ with $\sqrt{|J|+1}$). ◇

5. Applications and numerical experiments. In this section, we illustrate how to apply the theory presented in Section 4 to study the sensitivity of the LASSO solution to the tuning parameter λ when A is a random subgaussian matrix and when $m \ll n$. As already mentioned in the introduction, this case study is motivated by compressed sensing applications [2, 16, 18, 28, 45].

5.1. Application to LASSO parameter sensitivity. We start by recalling some standard notions from high-dimensional probability. For a general introduction to the topic we refer to, e.g., [44, 46].

Definition 5.1 (Subgaussian random variable and vector). *We call a real-valued random variable X K-subgaussian, for some $K > 0$, if $\|X\|_{\psi_2} \leqslant K$ where*

$$\|X\|_{\psi_2} := \inf\left\{t > 0 : \mathbb{E}\exp(X^2/t^2) \leqslant 2\right\}.$$

A real-valued random vector $Y \in \mathbb{R}^n$ is K-subgaussian if

$$\|Y\|_{\psi_2} := \sup_{\|z\|=1} \|\langle Y, z\rangle\|_{\psi_2} \leqslant K.$$

Definition 5.2 (Subgaussian matrix). *We call $A \in \mathbb{R}^{m\times n}$ a K-subgaussian matrix if the rows $A_i \in \mathbb{R}^n$ of A are independent subgaussian random vectors satisfying $\max_i \|A_i\|_{\psi_2} \leqslant K$ and $\mathbb{E}\, A_i A_i^T = I_n$ (where I_n is the $n \times n$ identity matrix). We call $\tilde{A} \in \mathbb{R}^{m\times n}$ a normalized K-subgaussian matrix if $\tilde{A} := m^{-1/2}A$ where $A \in \mathbb{R}^{m\times n}$ is a K-subgaussian matrix.*

The subgaussian random matrix model is quite popular in compressed sensing and it encompasses random matrices with independent and identically distributed Gaussian or Bernoulli entries. For more details, we refer to, e.g., [18, Chapter 9] or [44, Chapter 4]. It is well known that the singular values of subgaussian random matrices are well-behaved. This statement is formalized as Lemma B.1 in the supplement, and used to establish the following uniform control over the minimum singular values over all s-element sub-matrices A_I. Its proof is deferred to the supplement.

Proposition 5.3. *Suppose $\delta \in (0,1)$ and $A \in \mathbb{R}^{m\times n}$ is a normalized K-subgaussian matrix with $m \geqslant C\delta^{-2}K^2 \log K(s+1)\log(\mathrm{e}n/s)$ for some integer $1 \leqslant s < m$. Then, with probability at least $1 - \frac{2s}{\mathrm{e}n}$,*

$$\min_{|I|=s} \sigma_{\min}(A_I) \geqslant 1 - \delta.$$

Thus, observe that with the given choice of ε in the proof, one obtains a uniform lower bound to $\sigma_{\min}(A_I)$ independent of m, s, and n, when m is chosen large enough. Proposition 5.3 allows us to obtain an upper bound to the Lispchitz constant L in Corollary 4.11 that is independent of the support set I and holds with high probability when A is a normalized K-subgaussian matrix, provided that the LASSO solution is sparse enough and that Assumption 4.3 holds.

Proposition 5.4 (LASSO parameter sensitivity for subgaussian matrices under sparsity condition). *Let $A \in \mathbb{R}^{m\times n}$ be a normalized K-subgaussian matrix and fix an integer $1 \leqslant s < m$. Suppose $\delta \in (0,1)$ and $m \geqslant CK^2 \log K\delta^{-2} s \log(\mathrm{e}n/s)$ where $C > 0$ is a universal constant. The following holds with probability at least $1 - \frac{2s}{\mathrm{e}n}$ on the realization of A. If $\bar{b} \in \mathbb{R}^m$ and $\bar{\lambda} > 0$ are such that $\bar{x}(\bar{\lambda}) = S(\bar{\lambda})$ satisfies Assumption 4.3 and $\|\bar{x}(\bar{\lambda})\|_0 \leqslant s$, where S is the solution map in (4.1), then there exists $r > 0$ such that for all λ with $|\lambda - \bar{\lambda}| < r$,*

$$\|\bar{x}(\lambda) - \bar{x}(\bar{\lambda})\| \leqslant L|\lambda - \bar{\lambda}|, \qquad L < \frac{\sqrt{s+1}}{(1-\delta)^2}.$$

Proof. Let $I = I(\bar{\lambda}) := \mathrm{supp}(\bar{x}(\bar{\lambda}))$. By assumption, we have $|I| \leqslant s$. Via Proposition 5.3, restrict to the event where

$$\sigma_{\min}(A_{I(\bar{\lambda})}) \geqslant \min_{|J|=s} \sigma_{\min}(A_J) \geqslant 1 - \delta.$$

Since $\bar{x}(\bar{\lambda})$ satisfies Assumption 4.3, by Corollary 4.11 the solution mapping admits a locally Lipschitz localization about $\bar{\lambda}$, meaning there exists $r, L > 0$ such that for all $|\lambda - \bar{\lambda}| < r$ one has

$$\|\bar{x}(\lambda) - \bar{x}(\bar{\lambda})\| \leqslant L|\lambda - \bar{\lambda}|, \qquad L < \frac{\sqrt{|I|+1}}{\sigma_{\min}(A_I)^2} \leqslant \frac{\sqrt{s+1}}{(1-\delta)^2},$$

as desired. ■

Proposition 5.4 provides an upper bound to the Lispschitz constant of the solution map $\lambda \mapsto S(\lambda)$ of the LASSO at $\bar{\lambda}$ under Assumption 4.3 and provided that $\bar{x}(\bar{\lambda}) = S(\bar{\lambda})$ is sparse enough. We now show how to remove the sparsity condition when the measurements are of the form $\bar{b} = Ax + e$ for some (approximately) sparse vector x and bounded noise e and under a slightly stronger condition on the number of measurements m. To make this possible, a key ingredient is the recent analysis in [19] that provides explicit bounds to the sparsity of LASSO minimizers. We also need to introduce one more technical result from high-dimensional probability. Let $T_{n,s} := \sqrt{s}\mathbb{B}_1 \cap \mathbb{B}_2 \subset \mathbb{R}^n$. From [25, Corollary 1.2] and [44, §10.3] one may establish a deviation inequality result, Lemma B.2, which corresponds to estimating the *restricted isometry constants* of A (see, e.g., [18, Chapter 6] and references therein).

Proposition 5.5 (LASSO parameter sensitivity for subgaussian matrices). *Suppose that $x \in \mathbb{R}^n$ is s-sparse and that $\bar{b} := Ax + e$ where $A \in \mathbb{R}^{m\times n}$ is a normalized K-subgaussian matrix and $e \in \mathbb{R}^m$ with $\|e\| \leqslant \eta$ for some $\eta > 0$. For $\bar{\lambda} \geqslant 2\eta/\sqrt{s}$, define $\bar{x}(\bar{\lambda}) := S(\bar{\lambda})$, where S is the solution map in (4.1). Given $\delta < 0.7$ and $\varepsilon > 0$ assume that*

$$m \geqslant CK^2 \log K \delta^{-2} \left[(s+1)\log(\mathrm{e}n/s) + \log(2/\varepsilon)\right].$$

Then, with probability at least $1 - \frac{42s}{\mathrm{e}n} - \varepsilon$ on the realization of A the following holds: if Assumption 4.3 holds, then there exists $r > 0$ and a constant $C_\delta > 0$ depending only on δ such that for all $|\lambda - \bar{\lambda}| \leqslant r$,

$$\|\bar{x}(\lambda) - \bar{x}(\bar{\lambda})\| \leqslant L|\lambda - \bar{\lambda}|, \qquad L \leqslant C_\delta\sqrt{s}.$$

Proof. Let $\bar{x}(\bar{\lambda}) := S(\bar{\lambda})$ and denote $I := \operatorname{supp}(\bar{x}(\bar{\lambda}))$, $s_{\bar{\lambda}} := |I|$. Via Proposition 5.3, restrict to the event where

$$\sigma_{\min}(A_{I(\bar{\lambda})}) \geqslant \min_{|J|=s} \sigma_{\min}(A_J) \geqslant 1 - \delta.$$

Since $m \geqslant CK^2 \log K\delta^{-2}\left[s\log(\mathrm{e}n/s) + \log(2/\varepsilon)\right]$, by Lemma B.2 it holds with probability at least $1 - \varepsilon$ that

$$\sup_{x \in T_{n,s}} \left|\|Ax\| - \|x\|\right| \leqslant CK\sqrt{\log K} \cdot \frac{\sqrt{s\log(\mathrm{e}n/s)} + \sqrt{\log(2/\varepsilon)}}{\sqrt{m}} \leqslant \delta.$$

In particular, in this event, A satisfies the restricted isometry property of order s with constant δ (see, e.g., [18, Chapter 6]). Thus, under the assumption that $\delta < 0.7$ and the assumption on $\bar{\lambda}$, it holds by [19, Theorem 1.1] (cf. the discussion after [19, (3.7)]) that $s_{\bar{\lambda}} < 21s$.

Restricted to the joint event holding with probability at least $1 - \frac{2s_{\bar{\lambda}}}{en} - \varepsilon \geqslant 1 - \frac{42s}{en} - \varepsilon$, if Assumption 4.3 holds, then $\lambda \mapsto S(\lambda)$ admits a locally Lipschitz localization at $\bar{\lambda}$. Hence, there exists $r > 0$ such that for all λ satisfying $|\lambda - \bar{\lambda}| \leqslant r$ we have

$$\|\bar{x}(\lambda) - \bar{x}(\bar{\lambda})\| \leqslant L|\lambda - \bar{\lambda}|, \qquad L \leqslant \frac{\sqrt{1 + s_\lambda}}{\sigma^2_{\min}(A_I)} \leqslant \frac{\sqrt{1+21s}}{(1-\delta)^2} \leqslant C_\delta\sqrt{s}.$$ ■

Remark 5.6 (From sparsity to compressibility). Observe that the above result extends to s-compressible signals, i.e. signals $x \in \mathbb{R}^n$ for which, informally, the best s-term approximation error $\sigma_s(x)_1 := \inf\{\|x-z\|_1 : z \in \mathbb{R}^n, \|z\|_0 \leqslant s\}$ is small. This can be seen by letting $\bar{b} := Ax_s + e'$, where x_s is a best s-term approximation to x with respect to the ℓ_1-norm, i.e. an s-sparse vector such that $\|x - x_s\|_1 = \sigma_s(x)_1$, and $e' := A(x - x_s) + e$. We note, however, that this would require the knowledge of an upper bound to $\|A(x-x_s)\|$ in order to satisfy the assumption $\|e'\| \leqslant \eta$. ◇

5.2. Numerical experiments. We conclude this section by illustrating some numerical experiments. Suppose that $x \in \mathbb{R}^n$ is s-sparse and let $\bar{b} := Ax + \gamma w$ where $A \in \mathbb{R}^{m\times n}$ has $A_{ij} \overset{\text{iid}}{\sim} \mathcal{N}(0, 1/m)$, $\gamma = 0.1$ and $w_i \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. For $\lambda > 0$ define $\bar{x}(\lambda) := S(\lambda)$ where S is as given in (4.1), and define

$$\lambda^* := \inf \operatorname{argmin} \|\bar{x}(\lambda) - x\|. \tag{5.1}$$

Finally, let $I := \operatorname{supp}(\bar{x}(\lambda^*))$ and $s_{\lambda^*} := |I|$. In Figure 1 we plot $\|\bar{x}(\lambda) - \bar{x}(\lambda^*)\|$ as a function of λ (red curve) and superpose the Lipschitz upper bound evaluated at $\bar{\lambda} = \lambda^*$, namely $\sqrt{1+s_{\lambda^*}}\cdot\sigma_{\min}^{-2}(A_I)\cdot|\lambda-\lambda^*|$ (blue curve). Included on each plot is the ratio of the two quantities,

$$\frac{\sqrt{1+s_{\lambda^*}}\cdot\sigma_{\min}^{-2}(A_I)\cdot|\lambda-\lambda^*|}{\|\bar{x}(\lambda)-\bar{x}(\lambda^*)\|},$$

providing an alternative visualization of the extent of the bound's tightness in each setting. This latter curve is shown in purple, and it is plotted with respect to the purple axis appearing on the right-hand side of each plot.

The synthetic experiments were conducted for $s = 3, 7, 15$ (corresponding to each row in the figure, respectively) and $m = 50, 100, 150, 200$ (corresponding to each column from left to right, respectively) with $N = 200$. We selected $\gamma = 0.1$ and for each $j \in \operatorname{supp}(x) := \{1, \ldots, s\}$, $x_j = m + \sqrt{m}W$ where $W \sim \mathcal{N}(0, 1)$. Note that for each choice of (s, m), λ^* was chosen as the empirically best choice of tuning parameter from a logarithmically spaced grid of 501 λ values approximately centered about a nearly asymptotically order optimal choice $\gamma\sqrt{2\log n}$. The LASSO program was solved in Python using `lasso_path` from scikit-learn [39].

In Table 1 we report parameter values and relevant quantities associated with Figure 1. In particular, the upper bound on the Lipschitz constant is given by $L := \sqrt{1+s_{\lambda^*}}\cdot\sigma_{\min}^{-2}(A_I)$. The quantity in the penultimate column determines whether Assumption 4.3 holds (*n.b.*, A is a Gaussian random matrix, so A_I has full rank almost surely). For all values of s and m in the experiment, $\|A_{I^C}(\bar{b} - A_I\bar{x}_I)\|_\infty < \lambda^*$.

In every panel of the figure, the blue curve is indeed a strict (local) upper bound on $\|\bar{x}(\lambda)-\bar{x}(\lambda^*)\|$, supporting our theory. As predicted by the theory, the error $\|\bar{x}(\lambda)-\bar{x}(\lambda^*)\|$ is more pronounced when the problem is *under-regularized* ($\lambda < \lambda^*$), and when the support size grows. Note that the $(s, m) = (15, 50)$ plot (lower-left panel in the figure) corresponds to an unsuccessful approximate recovery of the ground truth signal x, due to the relatively large sparsity as compared to the number of measurements. In this setting, the small value

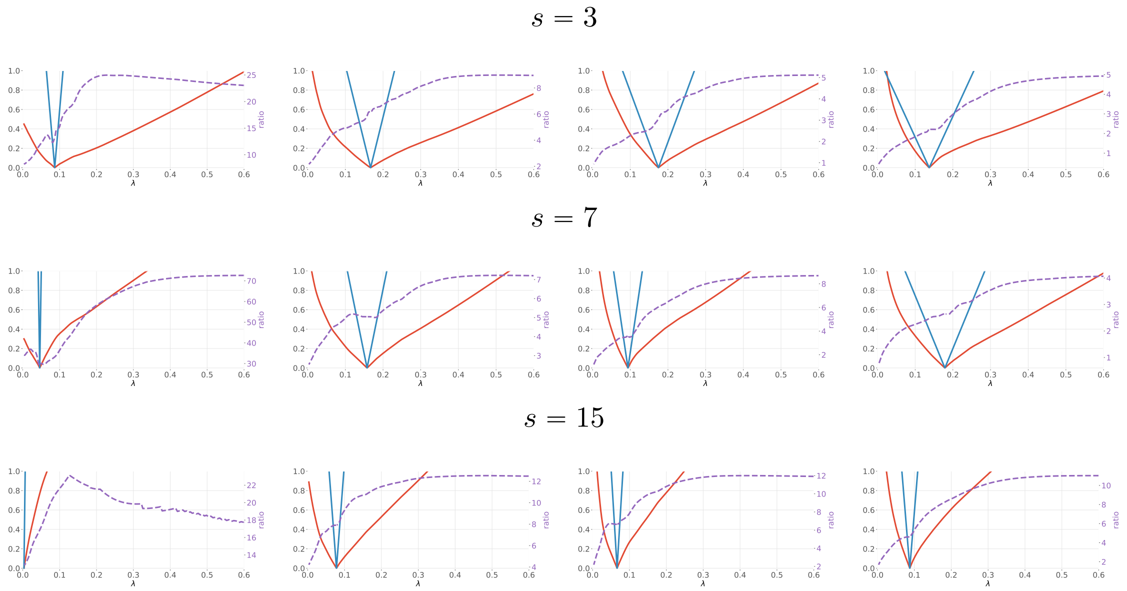

Figure 1: Lipschitzness of the solution mapping for λ about $\bar{\lambda} := \lambda^*$ as defined in (5.1). The red curve plots $\|\bar{x}(\lambda) - \bar{x}(\bar{\lambda})\|$; the blue curve, $|\lambda - \bar{\lambda}| \cdot \sqrt{1+s}/\sigma_{\min}(A_I)^2$. The ratio of the two is given by the purple curve, whose y-axis is on the right side of each plot. From left to right, each column corresponds to $m = 50, 100, 150, 200$, respectively. See Table 1 for parameter settings and variable values.

of $\sigma_{\min}(A_I)$ and the apparently poor behaviour of $\|\bar{x}(\lambda) - \bar{x}(\lambda^*)\|$ is consisent with our theory. Similarly, the $(s, m) = (7, 50)$ plot (middle-left panel in the figure) corresponds with a small value of $\sigma_{\min}(A_I)$ due to the relatively small value of m (and incidentally, it corresponds with poor recovery of the ground truth). Again, the relatively poor behaviour of $\|\bar{x}(\lambda) - \bar{x}(\lambda^*)\|$ (as compared, say, with $m \geqslant 100$) is consistent with our theory.

6. Final remarks.

In this paper we studied the optimal value and the optimal solution function of the LASSO problem as a function of both the right-hand side (or vector of measurements) $b \in \mathbb{R}^m$ and the regularization (or tuning) parameter $\lambda > 0$. Our analysis of the optimal value function is based on classical convex analysis, while the study of the optimal solution function is based on modern variational analysis (in particular, differentiation of set-valued maps). As a by-product we established the (strong) metric regularity of the subdifferential of the objective function at a solution. The assumptions needed to perform this analysis were inspired by uniqueness results in the literature, and are shown to hold, e.g., in the case where the right-hand side at the point in question admits an irreducible sparse representation with respect to the columns of the measurement matrix. We then combined these variational-analytic findings with random matrix-theoretic arguments to study the sensitivity of the LASSO solution with respect to the tuning parameter, providing upper bounds for the corresponding Lipschitz constant that hold with high probability for measurement matrices of subgaussian type.

s	m	N	η	s_{λ^*}	L	$\sigma_{\min}(A_I)$	$\lVert A_{I^C}(\bar{b} - A_I\bar{x}_I)\rVert_\infty$	λ^*
3	50	200	0.1	25	44.3	0.339	0.0831	0.0866
3	100	200	0.1	22	15.8	0.552	0.162	0.167
3	150	200	0.1	20	10.5	0.661	0.175	0.175
3	200	200	0.1	20	8.43	0.737	0.136	0.137
7	50	200	0.1	41	239	0.165	0.0458	0.046
7	100	200	0.1	26	19.1	0.522	0.155	0.158
7	150	200	0.1	43	26.2	0.503	0.0931	0.0939
7	200	200	0.1	23	9.44	0.72	0.175	0.179
15	50	200	0.1	63	303	0.162	0.00322	0.00326
15	100	200	0.1	47	50	0.372	0.078	0.0781
15	150	200	0.1	66	63	0.361	0.0655	0.0665
15	200	200	0.1	67	46.5	0.421	0.085	0.0876

Table 1: Parameter settings corresponding with Figure 1. The assumption $\lVert A_{I^C}(\bar{b} - A_I\bar{x}_I)\rVert_\infty < \lambda^*$ is satisfied for all entries where λ^* is defined as in (5.1).

Several questions arise naturally as a topic of future research. Can analogous statements to Theorem 4.9 be proved for alternative formulations of l_1 minimization such as the constrained LASSO, quadratically-constrained basis pursuit, or the square-root LASSO? Similarly, the extension of our analysis to convex regularizers beyond the l_1-norm is an interesting open issue. In particular, can the analysis carried out here be generalized to the matrix setting with the nuclear norm in place of the l_1-norm? This will certainly require a good handle on the graph of the subdifferential of the nuclear norm. Finally, Example 4.12 suggests that the Lipschitz bound with respect to λ provided by Corollary 4.11 might be hard to improve in general. Regarding applications to compressed sensing, another open problem is understanding whether our analysis could lead to results pertaining to the robustness to noise in the measurements. In the subgaussian case, the presence of $\sigma_{\max}(A)$ in the Lipschitz constant bound in Theorem 4.9 would lead to an algebraic dependence of the Lipschitz constant on the ambient dimension n. Therefore, whether it is possible to have a Lipschitz bound in b that does not depend on $\sigma_{\max}(A)$ remains an interesting open problem.

REFERENCES

[1] B. Adcock, A. Bao, and S. Brugiapaglia: *Correcting for unknown errors in sparse high-dimensional function approximation.* Numerische Mathematik 142(3), 2019, pp. 667–711.

[2] B. Adcock and A.C. Hansen: *Compressive Imaging: Structure, Sampling, Learning.* Cambridge University Press, Cambridge, UK, 2021.

[3] F.J. Aragón Artacho and M.H. Geoffroy: *Characterization of metric regularity of subdifferentials.* Journal of Convex Analysis 15(2), 2008, pp. 365–380.

[4] M. Bayati and A. Montanari: *The Lasso risk for Gaussian matrices.* IEEE Transactions on Infor-

mation Theory 58(4), 2011, pp. 1997–2017.
[5] A. Beck: *First-Order Methods in Optimization.* MOS-SIAM Series on Optimization, SIAM, Philadelphia, USA, 2017.
[6] A. Berk, Y. Plan, and Ö. Yilmaz: *Sensitivity of ℓ_1 minimization to parameter choice.* Information and Inference: A Journal of the IMA 10(2), 2021, pp. 397–453.
[7] A. Berk, Y. Plan, and Ö. Yilmaz: *On the best choice of LASSO program given data parameters.* IEEE Transactions on Information Theory 68(4), 2022, pp. 2573–2603.
[8] A. Beck and M. Teboulle: *A fast iterative shrinkage-thresholding algorithm for linear inverse problems.* SIAM Journal of Imaging Sciences 2(1), 2009, pp. 183–202.
[9] P.J. Bickel, Y. Ritov, and A.B. Tsybakov: *Simultaneous analysis of Lasso and Dantzig selector.* The Annals of Statistics, 37(4), 2009, pp. 1705–1732.
[10] J.F. Bonnans and A. Shapiro: *Perturbation Analysis of Optimization Problems.* Springer Series in Operations Research. Springer, New York, 2000.
[11] E.J. Candès, J. Romberg, and T. Tao: *Stable signal recovery from incomplete and inaccurate measurements.* Communications on Pure and Applied Mathematics 59, 2006, pp. 1207–1223.
[12] E.J. Candès, and T. Tao: *Decoding by linear programming.* IEEE Transactions on Information Theory 51(12), 2005, pp. 4203–4215.
[13] S.S. Chen, D.L. Donoho, and M.A. Saunders: *Atomic decomposition by basis pursuit.* SIAM Journal on Scientific Computing 2(1), 1998, pp. 33–61.
[14] D.L. Donoho: *Compressed sensing* IEEE Transactions on information theory 52(4), 2006, pp. 1289–1306.
[15] A.L. Dontchev and R.T. Rockafellar: *Implicit Functions and Solution Mappings: a View from Variational Analysis, 2nd Edition.* Springer Series in Operations Research and Financial Engineering. Springer, New York, Heidelberg, Dordrecht, London, 2014.
[16] Y.C. Eldar and G. Kutyniok, eds: *Compressed Sensing: Theory and Applications.* Cambridge University Press, Cambridge, UK, 2012.
[17] M.P. Friedlander, A. Goodwin, and T. Hoheisel: *From perspective maps to epigraphical projections.* arXiv:2102.06809, 2021.
[18] S. Foucart and H. Rauhut: *A Mathematical Introduction to Compressive Sensing.* Applied and Numerical Harmonic Analysis, Birkhäuser, New York, NY, 2013.
[19] S. Foucart, E. Tadmor, and M. Zhong: *On the sparsity of LASSO minimizers in sparse data recovery.* arXiv:2004.04348, 2020.
[20] J.-J. Fuchs: *On sparse representations in arbitrary redundant bases.* IEEE Transactions on Information Theory 50(6), 2004, pp. 1341–1344.
[21] T. Hastie, R. Tibshirani, and M. Wainwright: *Statistical Learning with Sparsity.* Chapman and Hall/CRC, New York, NY, 2015.
[22] J.C. Gilbert: *On the solution uniqueness characterization in the L1 norm and polyhderal gauge recovery.* Journal of Optimization Theory and Applications 172, 2017, pp. 70–101.
[23] T. Hoheisel: *Topics in Convex Analysis in Matrix Space.* Lecture Notes, Spring School on Variational Analysis, Paseky nad Jizerou, Czech Republic, 2019.
[24] R. Horn and C. R. Johnson: *Matrix Analysis.* Cambridge University Press, Cambridge, 2nd Edition, 2013.
[25] H. Jeong, X. Li, Y. Plan, and Ö. Yilmaz: *Sub-gaussian matrices on sets: optimal tail dependence and applications.* arXiv:2001.10631, 2020.
[26] C. Kanzow and T. Lechner: *Globalized inexact proximal Newton-type methods for nonconvex composite functions.* Computational Optimization and Applications 78, pp. 377–410.
[27] P.D. Khanh, B.S. Mordukhovich, and V.T. Phat: *A generalized Newton method for subgradient systems.* arXiv:2009.10551, 2021.
[28] M.-J. Lai and Y. Wang: *Sparse Solutions of Underdetermined Linear Systems and Their Applications.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2021.
[29] J.D. Lee, Y. Sun, and M.A. Saunders: *Proximal Newton-type methods for minimizing composite functions.* SIAM Journal on Optimization 24(3), 2014, pp. 1420–1443.
[30] A. Milzarek and M. Ulbrich: *A semismooth Newton method with multidimensional filter global-*

ization for L1-optimization. SIAM Journal on Optimization 24, 2014, pp. 298–333.

[31] B.S. Mordukhovich: *Variational Analysis and Generalized Differentiation. I: Basic Theory.* Grundlehren Series (Fundamental Principles of Mathematical Sciences), Vol. 330, Springer, Berlin, 2006.

[32] B.S. Mordukhovich: *Variational Analysis and Applications.* Springer Monographs in Mathematics book series, Springer International Publishing AG, 2018.

[33] B. K. Natarajan: *Sparse approximate solutions to linear systems.* SIAM Journal on Computing 24, 1995, pp. 227–234.

[34] S. Oymak, C. Thrampoulidis, and B. Hassibi: *The squared-error of generalized Lasso: a precise analysis.* In 2013 51st Annual Allerton Conference on Communication, Control, and Computing (Allerton), 2013, pp. 1002–1009.

[35] T. Park and G. Casella: *The Bayesian Lasso.* Journal of the American Statistical Association 103(482), 2008, pp. 681–686.

[36] R.T. Rockafellar: *Convex Analysis.* Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, NJ, 1970.

[37] R.T. Rockafellar and R.J.-B. Wets: *Variational Analysis.* Grundlehren der Mathematischen Wissenschaften, Vol. 317, Springer-Verlag, Berlin, 1998.

[38] Y. Shen, B. Han, and E. Braverman: *Stable recovery of analysis based approaches.* Applied and Computational Harmonic Analysis, 39(1), 2015, pp. 161–172.

[39] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, *et al.*: *Scikit-learn: Machine learning in Python.* The Journal of machine Learning research, 12, 2011, pp. 2825–2830.

[40] C. Thrampoulidis, A. Panahi, and B. Hassibi: *Asymptotically exact error analysis for the generalized ℓ_2^2-Lasso.* In 2015 IEEE International Symposium on Information Theory (ISIT), 2015, pp. 2021–2025.

[41] C. Thrampoulidis, E. Abbasi, and B. Hassibi: *Precise error analysis of regularized M-estimators in high dimensions.* IEEE Transactions on Information Theory, 64(8), 2018, pp. 5592–5628.

[42] R. Tibshirani: *Regression shrinkage and selection via the lasso.* Journal of the Royal Statistic Society (Series B) 58, 1996, pp. 267–288.

[43] R.J. Tibshirani: *The lasso problem and uniqueness.* Electronic Journal of Statistics 7, 2013, pp. 1456–1490.

[44] R. Vershynin: *High-Dimensional Probability: an Introduction with Applications in Data Science.* Cambridge University Press, Cambridge, UK, 2017.

[45] M. Vidyasagar: *An introduction to compressed sensing.* Society for Industrial and Applied Mathematics, Philadelphia, NY, 2019.

[46] M.J. Wainwright: *High-dimensional statistics: A non-asymptotic viewpoint.* Cambridge University Press, Cambridge, UK, 2019.

[47] H. Zhang, W. Yin, and L. Cheng: *Necessary and sufficient conditions of solution uniqueness in 1-norm minimization.* Journal of Optimization Theory and Applications 164, 2015, pp. 109–122.

Appendix A. Supplement to Section 4. In this section provide further details on the results for the solution map of LASSO. We begin with a proof of Lemma 4.8.

Proof of Lemma 4.8. Consider the optimization problem

$$\max_{\binom{v}{\alpha}\in\mathbb{R}^{m+1}} \phi(v,\alpha) := \left\| [I\ d]\begin{pmatrix} v \\ \alpha \end{pmatrix} \right\|^2 \text{ s.t. } \left\| \begin{pmatrix} v \\ \alpha \end{pmatrix} \right\|^2 = 1.$$

This problem (by compactness) has a solution, which then necessarily satisfies the Karush-

Kuhn-Tucker (KKT) conditions

$$v + \alpha d = \tau v, \tag{A.1}$$

$$d^T v + \alpha \|d\|^2 = \tau \alpha, \tag{A.2}$$

$$\|v\|^2 + \alpha^2 = 1. \tag{A.3}$$

It is clear that $\tau \neq 0$ (since otherwise $(v, \alpha) = 0$ which is not feasible). We now assume first that $d = 0$. In this case $\alpha = 0$ (by (A.2) and $\tau \neq 0$). Hence v can be any normed vector, which leads to an objective value 1.

Now assume that $d \neq 0$. Setting $\hat{\alpha} := 0$, (A.2) reads $v^T d = 0$. Set $\hat{\tau} := 1$ and choose any normed $\hat{v} \in \{d\}^\perp$. Then $(\hat{v}, \hat{\alpha}, \hat{\tau})$ satisfies the KKT conditions with objective value $\phi(\hat{v}, \hat{\alpha}) = 1$. On the other hand, for $\alpha \neq 0$, we first observe that

$$\tau d^T v = d^T v + \alpha \|d\|^2 = \tau \alpha,$$

where the first identity uses (A.1) and the second one uses (A.2). As $\tau \neq 0$, this yields $\alpha = d^T v$, and hence (A.2) gives $\tau\alpha = \alpha(1 + \|d\|^2)$. Since $\alpha \neq 0$, this shows that $\tau = 1 + \|d\|^2$. Inserting in (A.1) and using that $d \neq 0$ now yields

$$v = \frac{\alpha}{\|d\|^2} d. \tag{A.4}$$

In particular, $\|v\| = \frac{\alpha}{\|d\|}$. Thus, (A.3) implies $1 - \alpha^2 = \|v\|^2 = \frac{\alpha^2}{\|d\|^2}$, and consequently we have $\alpha = \frac{\|d\|}{\sqrt{1+\|d\|^2}}$. Therefore, (A.4) yields $v = \frac{1}{\sqrt{1+\|d\|^2}} \cdot \frac{d}{\|d\|}$. All in all, we find that $(\bar{v}, \bar{\alpha}, \bar{\tau})$ with

$$\bar{v} := \frac{1}{\sqrt{1 + \|d\|^2}} \cdot \frac{d}{\|d\|}, \quad \bar{\alpha} := \frac{\|d\|}{\sqrt{1 + \|d\|^2}}, \quad \bar{\tau} := 1 + \|d\|^2$$

is a KKT triple with objective value $\phi(\bar{v}, \bar{\alpha}) = 1 + \|d\|^2 > 1 = \phi(\hat{v}, \hat{\alpha})$. Since no other KKT points exist when $d \neq 0$, this concludes the proof. ■

Next, we provide detailed calculations for Remark 4.10, in which A is treated as a parameter of the solution map.

Remark A.1 (The matrix A as a parameter). Theorem 4.9 can be extended to the case where the solution is considered as a function of (b, λ, A). With the (extended) function $f(A, b, \lambda, z) = \frac{1}{\lambda}\left(A^T(Ax - b)\right)$, a direct computation shows that $D_A f(\bar{A}, \bar{b}, \bar{\lambda}, \bar{x})H = \bar{\lambda}^{-1}((\bar{A}^T H + H^T \bar{A})\bar{x} - H^T \bar{b})$, for any $H \in \mathbb{R}^{m \times n}$ (where, in this case, $\mathbb{R}^{m \times n}$ has been equipped with the Frobenius norm $\|\cdot\|_F$). This yields

$$\begin{aligned}
&\left\|D_{(A,b,\lambda)} f(\bar{A}, \bar{b}, \bar{\lambda}, \bar{x})\right\| \\
&\qquad = \frac{1}{\bar{\lambda}} \max_{\|(H,v,\alpha)\| \leqslant 1} \left\|(\bar{A}^T H + H^T \bar{A})\bar{x} - H^T \bar{b} - \bar{A}^T v - \alpha \bar{\lambda}^{-1}\left(\bar{A}^T(\bar{A}\bar{x} - \bar{b})\right)\right\| \\
&\qquad \leqslant \frac{1}{\bar{\lambda}} \max_{\|H\|_F \leqslant 1} \left\|(\bar{A}^T H + H^T \bar{A})\bar{x} - H^T \bar{b}\right\| + \frac{1}{\bar{\lambda}} \max_{\|(v,\alpha)\| \leqslant 1} \left\|\bar{A}^T v + \alpha \bar{\lambda}^{-1}\left(\bar{A}^T(\bar{A}\bar{x} - \bar{b})\right)\right\|.
\end{aligned}$$

Note that the second term above was estimated in the proof of Theorem 4.9. Hence, we focus on the first term. Observing that $\|H\| \leqslant \|H\|_F \leqslant 1$, we see that

$$\max_{\|H\|_F=1} \left\|(\bar{A}^T H + H^T \bar{A})\bar{x} - H^T \bar{b}\right\| \leqslant \max_{\|H\|_F=1} \left\|\bar{A}^T H\bar{x}\right\| + \left\|H^T(\bar{A}\bar{x} - \bar{b})\right\| \leqslant \|\bar{A}\|\|\bar{x}\| + \|\bar{A}\bar{x} - \bar{b}\|.$$

Combining the above inequalities and using an argument analogous to the proof of Theorem 4.9, we obtain that the solution map $(A, b, \lambda) \mapsto \operatorname{argmin}_{z\in\mathbb{R}^n} \left\{\frac{1}{2}\|Az - b\|^2 + \lambda\|z\|_1\right\}$ is locally Lipschitz at $(\bar{A}, \bar{b}, \bar{\lambda})$ with (local) Lipschitz modulus L satisfying

$$L \leqslant \frac{1}{\sigma_{\min}(\bar{A}_I)^2} \left(\left\|\bar{A}\bar{x} - \bar{b}\right\| + \sigma_{\max}(\bar{A}) \left(\|\bar{x}\| + \sqrt{1 + \left\| \frac{\bar{A}\bar{x} - \bar{b}}{\bar{\lambda}} \right\|^2} \right) \right).$$

$\diamond$

Appendix B. Supplement to Section 5. We first give a precise statement of the sense in which singular values of subgaussian matrices are well-behaved. Combining Lemma B.1 with a union bound is the key to Proposition 5.3.

Lemma B.1 (Spectral bound for K-subgaussian matrices). *Let $A \in \mathbb{R}^{m\times n}$ have rows A_i that are independent, mean-zero subgaussian isotropic random vectors in $\mathbb{R}^n$. For any $t \geqslant 0$ we have*

$$\sqrt{m} - CK\sqrt{\log K}(\sqrt{n} + t) \leqslant \sigma_{\min}(A) \leqslant \sigma_{\max}(A) \leqslant \sqrt{m} + CK\sqrt{\log K}(\sqrt{n} + t)$$

with probability at least $1 - 2\exp(-t^2)$ where $K := \max_i \|A_i\|_{\psi_2}$.

See [44, Theorem 4.6.1] for a proof of this result. However, note that we have stated a version that uses the optimal dependence on K established in [25]. Next, we give a proof of Proposition 5.3.

Proof of Proposition 5.3. Applying Lemma B.1, for any single s-element subset I of the columns we have, with probability at least $1 - 2\exp(-t^2)$,

$$\sigma_{\min}(A_I) \geqslant 1 - \frac{CK\sqrt{\log K}(\sqrt{s} + t)}{\sqrt{m}}.$$

Union bounding over all $\binom{n}{s}$ s-element subsets I gives

$$\min_{|I|=s} \sigma_{\min}(A_I) \geqslant 1 - \frac{CK\sqrt{\log K}(\sqrt{s} + t)}{\sqrt{m}}$$

with probability at least $1 - 2\binom{n}{s}\exp(-t^2)$. Using Stirling's approximation we see that the probability of failure is bounded by

$$2\binom{n}{s}\exp(-t^2) \leqslant 2\left(\frac{en}{s}\right)^s \exp(-t^2).$$

The right-hand side above is less than or equal to ε if

$$t \geqslant \sqrt{s \log\left(\frac{\mathrm{e}n}{s}\right) + \log\left(\frac{2}{\varepsilon}\right)}.$$

Choosing the probability of failure to be $\varepsilon = 2s/(\mathrm{e}n)$ yields the desideratum with an appropriate choice of universal constant $C > 0$. ■

Finally, we include a technical lemma from high-dimensional probability, which is required to prove Proposition 5.5.

Lemma B.2 (Restricted isometry constants of a subgaussian matrix). *Let $A \in \mathbb{R}^{m\times n}$ be a K-subgaussian matrix. Then,*

$$\mathbb{E} \sup_{x \in T_{n,s}} \left| \|Ax\| - \sqrt{m}\|x\| \right| \leqslant CK\sqrt{\log K}\left[\sqrt{s \log(\mathrm{e}n/s)} + 1\right].$$

Moreover, for any $u \geqslant 0$, with probability at least $1 - 3\exp(-u^2)$,

$$\sup_{x \in T_{n,s}} \left| \|Ax\| - \sqrt{m}\|x\| \right| \leqslant CK\sqrt{\log K}\left[\sqrt{s \log(\mathrm{e}n/s)} + u\right].$$