LIPSCHITZ STABILITY OF LEAST-SQUARES PROBLEMS REGULARIZED BY FUNCTIONS WITH $\mathcal{C}^2$-CONE REDUCIBLE CONJUGATES

YING CUI, TIM HOHEISEL, TRAN T. A. NGHIA, AND DEFENG SUN

ABSTRACT. In this paper, we study Lipschitz continuity of the solution mappings of regularized least-squares problems for which the convex regularizers have (Fenchel) conjugates that are $\mathcal{C}^2$-cone reducible. Our approach, by using Robinson's strong regularity on the dual problem, allows us to obtain new characterizations of Lipschitz stability that rely solely on first-order information, thus bypassing the need to explore second-order information (curvature) of the regularizer. We show that these solution mappings are automatically Lipschitz continuous around the points in question whenever they are locally single-valued. We leverage our findings to obtain new characterizations of full stability and tilt stability for a broader class of convex additive-composite problems.

Keywords. Least-squares, Lipschitz stability, full stability, tilt stability, convex optimization, Lasso, variational analysis and nonsmooth optimization, second-order analysis, regularized problem.

Mathematics Subject Classification (2020). 49J52, 49J53, 49K40, 90C25, 90C31

1. INTRODUCTION

Stability (or sensitivity) analysis in optimization is the study of how the optimal solutions (or the optimal value) of an optimization problem varies as the problem-defining data (parameters) changes. This field has a rich and longstanding tradition anchored by the pioneering works of Kojima [17, 18] and Robinson [28, 29, 30], primarily focused on smooth, nonlinear programming in finite dimensions. The monograph by Bonnans and Shapiro [2] presents the state-of-the-art treatment for problems with smooth data, in particular, in infinite dimensions. A rich source of general theory for nonsmooth problems, based on graphical differentiation of set-valued maps, is provided in the monograph by Dontchev and Rockafellar [9].

Recently, there has been renewed interest in stability analysis for the following nonsmooth, convex optimization problems which occur ubiquitously in machine learning and statistical estimations:

$$\min_{x \in \mathbb{X}} \quad \frac{1}{2\mu}\|Ax - b\|^2 + g(x), \tag{1}$$

where $A : \mathbb{X} \to \mathbb{Y}$ is a linear operator between two Euclidean spaces $\mathbb{X}$ and $\mathbb{Y}$, b is a vector in $\mathbb{Y}$, $\mu > 0$ is the tuning parameter, and $g : \mathbb{X} \to \mathbb{R} \cup \{+\infty\}$ is a closed, proper, convex function (usually called a regularizer in machine learning). When the triple (A, b, μ) are considered parameters, it is important to study stability of the solution mapping of problem (1) defined by

$$S(A, b, \mu) := \operatorname*{argmin}_{x \in \mathbb{X}} \left\{ \frac{1}{2\mu}\|Ax - b\|^2 + g(x) \right\}. \tag{2}$$

Specifically, in this paper, we aim to determine the conditions under which the solution mapping (2) is single-valued and Lipschitz continuous when there are small perturbations in the initial data (A, b, μ) of problem (1).

To the best of our knowledge, there are currently two main approaches for studying the stability of this solution mapping. The first one employs the toolkit of modern variational analysis à la Dontchev and Rockafellar [9]; see also the standard texts by Mordukhovich [20], or Rockafellar and

Wets [27]. Some recent contributions along these lines were made by Berk et al. [3, 4], Gfrerer and Outrata [14], Hu et al. [15] (which strengthens the qualitative but not the quantitative results of Berk et al.), Meng et al. [22], Nghia [24], and Vaiter et al. [32]. When the regularizer g is the ℓ_1 norm, the problem (1) is known as the Lasso (Least Absolute Shrinkage and Selection Operator) problem. Berk et al. [3] provide a sufficient condition [3, Assumption 4.3] at which the solution mapping S is single-valued, Lipschitz continuous and directionally differentiable around a fixed triple $(\bar{A}, \bar{b}, \bar{\mu})$. This condition appeared in [23, 34] as a sufficient condition for the uniqueness of a given solution to the Lasso problem. The approach in [3] relies on the well-known Mordukhovich criterion [20, Theorem 4.18] for *metric regularity* and available second-order information of the ℓ_1 norm. On the other hand, in the recent paper [24], Nghia proposed a different approach via Robinson's *generalized implicit function theorem* [28] and *tilt stability* [25] for more general functions g, including popular regularizers such as the ℓ_1 norm, the ℓ_1/ℓ_2 norm, and the nuclear norm. It is revealed in [24, Theorem 3.7] that the aforementioned sufficient condition in [3] is also necessary for the Lipschitz stability of the solution mapping (2) for the Lasso problem. Some recent papers [15, 22] obtain impressive results about Lipschitz continuity of the set-valued solution mapping (2) with fixed operator A *relative* to its domain when g is the ℓ_1 norm. Their Lipschitz stability is different from ours in this paper and their approaches strongly rely on the polyhedral structure of the ℓ_1 norm and its subdifferential mapping. When regularizer g is *partly smooth*, Vaiter et al [32, Theorem 1] proves a strong result that $S(\bar{A}, \cdot, \bar{\mu})$ is continuously differentiable around $\bar{b}$ when $\bar{x} = S(\bar{A}, \bar{b}, \bar{\mu})$ is a *strong minimizer* of the corresponding problem, but they have to assume that $\bar{b}$ does not belong to a *transition space*. The latter condition seems to be involved and recently [24, Example 3.14] provides a simple example at which the solution mapping S is Lipschitz continuous, but it is not continuously differentiable.

The second approach, spearheaded by Bolte et al. [6], is to rewrite the optimality conditions as a (locally) Lipschitz equation, say, by using proximal operators, and to employ the machinery of implicit differentiation with *conservative Jacobians* [5]. The conservative Jacobians here serve as a versatile tool which, among other things, bypasses the failure of the chain rule for Clarke Jacobians. The ideas mentioned here can be extended to (maximally) monotone inclusions (which are not necessarily optimality conditions of a convex optimization problem) [7].

In this paper, we mainly study Lipschitz stability of solution mapping (2) with a fresh approach via *strong regularity* [28] via the dual problem. Strong regularity was introduced by Robinson in the landmark paper [28] for *generalized equations* with applications to *nonlinear programming* to study Lipschitz stability of the *Lagrange system*. Robinson's strong regularity can be characterized by the well-known *Linear Independence Constraint Qualification* and the *Strong Second-Order Sufficient Condition* for nonlinear programming; see, e.g., [28, 2, 16, 9] for more historic discussions and its immense applications to optimization theory and algorithms. In [21, Theorem 5.6], Mordukhovich, Nghia and Rockafellar give full characterizations of strong regularity for more general parametric constrained optimization problems in the following setting:

$$\min_{u \in \mathbb{U}} \quad \varphi(u, p) \quad \text{subject to} \quad G(u, p) \in \Theta, \tag{3}$$

with Euclidean spaces $\mathbb{U}, \mathbb{V}, \mathbb{P}$ ($\mathbb{P}$ being the parameter space), a closed and convex set $\Theta \subset \mathbb{V}$, and twice continuously differentiable cost function $\varphi : \mathbb{U} \times \mathbb{P} \to \mathbb{R}$ and constrained mapping $G : \mathbb{U} \times \mathbb{P} \to \mathbb{V}$. It is worth noting here that [21] crucially assumes that the constrained set Θ is $\mathcal{C}^2$-*reducible* in the sense of [2, Definition 3.135] that allows them to cover some previous works on the strong regularity for non-polyhedral settings such as the *second-order cone programming* [1] and the *semidefinite programming* [33]. Their characterizations are based on the so-called *Partial Constraint Nondegeneracy* and the *Generalized Strong Second-Order Sufficient Condition* that will be recalled in our Subsection 2.3.

A simple, but important observation of our paper is that the solution mapping (2) is, indeed, the mapping of (partial) Lagrange multipliers of the following constrained problem

$$\min_{(y,t)\in\mathbb{Y}\times\mathbb{R}} \quad \frac{\mu}{2}\|y\|^2 - \langle b,\, y \rangle + t \quad \text{subject to} \quad (A^*y, t) \in \text{epi}\, g^*, \tag{4}$$

where $\text{epi}\, g^*$ stands for the epigraph of the Fenchel conjugate function of g. This problem is in the format of (3) and equivalent to the Fenchel-Rockafellar dual of problem (1)

$$\min_{y\in\mathbb{Y}} \quad \frac{\mu}{2}\|y\|^2 - \langle b,\, y \rangle + g^*(A^*y).$$

This unveils the connection between the strong regularity of problem (4) and the Lipschitz stability of solution mapping (2) of problem (1). As mentioned above, in order to apply [21, Theorem 5.6], we need to suppose that the constrained set of problem (4), i.e, the epigraph of the conjugate function g^*, is $\mathcal{C}^2$-*cone reducible*. This assumption is not restrictive: it automatically holds for many important regularizers in optimization such as the ℓ_1 norm, the ℓ_1/ℓ_2 norm, the nuclear norm, and many more spectral functions [10] and support functions; see also our Remark 3.6.

While applying the theory of strong regularity in [21] to the dual problem (4), we observe that the Generalized Second-Order Sufficient Condition holds automatically and the Partial Constraint Nondegeneracy can be simplified with first-order information on the regularizer g. This significantly distinguishes our paper from [3, 24, 32] that directly work on the primal problem and needs more advanced *second-order* computation on the regularizer. In particular, we show that the solution mapping S is single-valued and Lipschitz continuous around a fixed triple $(\bar{A}, \bar{b}, \bar{\mu}) \in \mathcal{L}(\mathbb{X}, \mathbb{Y}) \times \mathbb{Y} \times \mathbb{R}_{++}$ of initial data for $\bar{x} \in S(\bar{A}, \bar{b}, \bar{\mu})$ if and only if

$$\ker \bar{A} \cap \text{par}\, \partial g^*(\bar{z}) = \{0\} \quad \text{with} \quad \bar{z} := -\frac{1}{\bar{\mu}}\bar{A}^*(\bar{A}\bar{x} - \bar{b}),$$

where $\text{par}\, \partial g^*(\bar{z})$ denotes the *parallel subspace* of the subdifferential set $\partial g^*(\bar{z})$. This characterization recovers some similar results obtained in [3, 24] for special cases of regularizers such as the ℓ_1 norm, the ℓ_1/ℓ_2 norm, and the nuclear norm. Although our technique relies on second-order variation analysis, our condition above does not need any second-order information of g.

Another key aspect of our paper studies characterizations of tilt stability [25] and full stability [19], both of which also relate to Lipschitz stability of the solution mapping of the following optimization problem

$$\min_{x\in\mathbb{X}} \quad f(x,p) + g(x), \tag{5}$$

where $f : \mathbb{X} \times \mathbb{P} \to \mathbb{R}$ is a twice continuously differentiable function that is convex with respect to $x \in \mathbb{X}$. This problem is more general than (1) at which the basic parameter p is considered as the triple parameter (A, b, μ). In our framework, full stability, as introduced by Poliquin, Levy, and Rockafellar [19], mainly analyzes Lipschitz continuity of the (local) optimal solution mapping of

$$\min_{x\in\mathbb{X}} \quad f(x,p) + g(x) - \langle v,\, x \rangle, \tag{6}$$

with respect to both basic parameter $p \in \mathbb{P}$ and tilt (linear) parameter $v \in \mathbb{X}$; while tilt stability is a particular case when the basic perturbation p is fixed. Both tilt stability and full stability are characterized by using different advanced nonsmooth second-order structures that may be difficult to compute [14, 19, 21, 24, 25]. In Section 4, by using our results on Lipschitz stability, we achieve a new pointwise characterization for both full stability and tilt stability of problem (5). Unlike other characterizations of full stability and tilt stability, our pointwise condition uses first-order information of the function g. Specifically, we show that $\bar{x} \in \mathbb{X}$ is a full (or tilt) stable optimal solution of problem (5) for parameter $\bar{p} \in \mathbb{P}$ if and only if $-\nabla_x f(\bar{x}, \bar{p}) \in \partial g(\bar{x})$ and

$$\ker \nabla^2_{xx} f(\bar{x}, \bar{p}) \cap \text{par}\, \partial g^*(-\nabla_x f(\bar{x}, \bar{p})) = \{0\};$$

see our Theorem 4.4 for further details. This condition still needs second-order information of the function f, but is easy to compute, as f is twice differentiable. The crucial part of this condition is that we do not need to calculate any second-order structures for the nonsmooth function g. Consequently, it is also a sufficient condition for Lipschitz stability of the solution mapping of problem (5). However, unlike the regularized least-squares problem (1), this condition is not necessary for the aforementioned Lipschitz stability of problem (5).

Notation: Throughout the paper, $\mathcal{L}(\mathbb{X}, \mathbb{Y})$ is the space of all linear operators from the Euclidean space $\mathbb{X}$ to the Euclidean space $\mathbb{Y}$. For $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$, we denote its *adjoint* by A^*, its *kernel* (or *null space*) by $\ker A$ and its *range* by $\operatorname{rge} A$. For a Fréchet differentiable map $G : \mathbb{X} \to \mathbb{Y}$, we write $\mathcal{J}G(x) \in \mathbb{Y} \times \mathbb{X}$ as the Jacobian matrix of G at $x \in \mathbb{X}$. When $\mathbb{Y} = \mathbb{R}$, we have $\mathcal{J}G(x) = \nabla G(x)^*$ for any $x \in \mathbb{X}$.

2. Preliminaries

Let $\mathbb{X}$ be a Euclidean space, i.e., a finite-dimensional real vector space with inner product denoted by $\langle \cdot, \cdot \rangle$. The (Euclidean) norm on $\mathbb{X}$ induced by the ambient inner product is denoted by $\|\cdot\|$. The closed ball in $\mathbb{X}$ with center $x \in \mathbb{X}$ and radius $r > 0$ is denoted by $\mathbb{B}_r(x)$.

2.1. **Tools from convex analysis.** For a function $f : \mathbb{X} \to \mathbb{R} \cup \{+\infty\}$, its *epigraph* is the set given by $\operatorname{epi} f := \{(x, \alpha) \in \mathbb{X} \times \mathbb{R} \mid f(x) \leqslant \alpha\}$ while its *domain* is the set $\operatorname{dom} f := \{x \in \mathbb{X} \mid f(x) < +\infty\}$. We call f *proper* if $\operatorname{dom} f$ is nonempty. We say that f is *convex* if $\operatorname{epi} f$ is convex, and we call it *closed* or *lower semicontinuous (lsc)* if $\operatorname{epi} f$ is closed. The *(Fenchel) conjugate* of a proper f is the function $f^* : X \to \mathbb{R} \cup \{+\infty\}$, given by $f^*(y) = \sup_{y \in \mathbb{X}} \{\langle y, x \rangle - f(x)\}$. Its *subdifferential* at $\bar{x} \in \mathbb{X}$ is the set $\partial f(\bar{x}) = \{y \in \mathbb{X} \mid f(\bar{x}) + \langle y, x - \bar{x} \rangle \leqslant f(x) \ \forall x \in \mathbb{X}\}$. When f is closed, proper, convex, throughout the paper we use the following relation frequently:

$$v \in \partial f(x) \iff x \in \partial f^*(v).$$

For a convex set $\Omega \subset \mathbb{X}$, its *relative interior* is given by

$$\operatorname{ri} \Omega := \{x \in \Omega \mid \exists \epsilon > 0 : \ \mathbb{B}_\epsilon(x) \cap \operatorname{aff} \Omega \subset \Omega\}, \tag{7}$$

where $\operatorname{aff} \Omega$ is the *affine hull* of Ω. The *conical hull* of Ω is $\operatorname{cone} \Omega := \mathbb{R}_+ \Omega$, the smallest (convex) cone containing Ω. The *span* of Ω is denoted $\operatorname{span} \Omega$. The *parallel subspace* of Ω is defined, for any $x \in \Omega$, by

$$\operatorname{par} \Omega := \operatorname{span}(\Omega - x). \tag{8}$$

Note that $\operatorname{par} \Omega = \operatorname{aff} \Omega - \operatorname{aff} \Omega = \operatorname{aff} \Omega - x$ for any $x \in \Omega$. Thus for any $x \in \operatorname{ri} \Omega$, we have

$$\operatorname{par} \Omega = \operatorname{cone}(\Omega - x). \tag{9}$$

The *lineality space* of Ω, denoted by $\operatorname{lin} \Omega$, is the largest subspace in Ω. Moreover, the *horizon cone* of Ω is

$$\Omega^\infty := \{v \in \mathbb{X} \mid \exists \{x_k \in \Omega\}, \{t_k\} \downarrow 0 : \ t_k x_k \to v\}, \tag{10}$$

which is a closed cone, and convex as Ω is. In particular, for a convex function f and $\bar{y} \in \partial f(\bar{x})$, we have

$$\partial f(\bar{x})^\infty \subset \operatorname{cone}(\partial f(\bar{x}) - \bar{y}), \tag{11}$$

cf. [27, Theorem 3.6]. The *normal cone* of Ω at $\bar{x}$ is

$$N_\Omega(\bar{x}) = \{y \in \mathbb{X} \mid \langle y, x - \bar{x} \rangle \leqslant 0 \ \forall x \in \Omega\}, \tag{12}$$

which is the subdifferential of the (convex) *indicator function* δ_Ω of Ω at $\bar{x}$, which is defined by $\delta_\Omega(x) = 0$ if $x \in \Omega$ and $\delta_\Omega(x) = +\infty$ otherwise. The following result, which is useful to our study, links many of the above mentioned objects.

Proposition 2.1. *Let $f : \mathbb{X} \to \mathbb{R} \cup \{+\infty\}$ be a closed, proper, convex function and $\bar{x} \in \mathbb{X}$ such that $\partial f(\bar{x}) \neq \varnothing$. Then*

$$\mathbb{X} \times \{0\} \cap \operatorname{span} N_{\operatorname{epi} f}(\bar{x}, f(\bar{x})) = \operatorname{par} \partial f(\bar{x}) \times \{0\}.$$

Proof. First, observe by [27, Theorem 8.9] that

$$N_{\operatorname{epi} f}(\bar{x}, f(\bar{x})) = \{\sigma(v, -1) \mid v \in \partial f(\bar{x}), \ \sigma > 0\} \bigcup \partial f(\bar{x})^{\infty} \times \{0\}. \tag{13}$$

Now, let $(v, 0) \in \mathbb{X} \times \{0\} \cap \operatorname{span} N_{\operatorname{epi} f}(\bar{x}, f(\bar{x}))$. In view of (13), we find scalars $\lambda_1, \dots, \lambda_r \in \mathbb{R}$, $\mu_1, \dots, \mu_s \in \mathbb{R}$, and vectors $v_1, \dots, v_r \in \partial f(\bar{x})$, $w_1, \dots, w_s \in \partial f(\bar{v})^{\infty}$ such that

$$(v, 0) = \sum_{i=1}^{r} \lambda_i(v_i, -1) + \sum_{j=1}^{s} \mu_j(w_j, 0).$$

It follows that $\sum_{i=1}^{r} \lambda_i = 0$, i.e., $\lambda_r = -\sum_{i=1}^{r-1} \lambda_i$. With the previous identity, we thus obtain

$$v = \sum_{i=1}^{r-1} \lambda_i(v_i - v_r) + \sum_{j=1}^{s} \mu_j w_j \subset \operatorname{par} \partial f(\bar{x}) + \operatorname{span} \partial f(\bar{x})^{\infty}.$$

For any $\bar{y} \in \partial f(\bar{x})$, note from (11) that

$$\operatorname{span} \partial f(\bar{x})^{\infty} \subset \operatorname{span}(\partial f(\bar{x}) - \bar{y}) = \operatorname{par} \partial f(\bar{x}).$$

Therefore, $v \in \operatorname{par} \partial f(\bar{x})$, which shows the desired inclusion.

Conversely, take $(v, 0) \in \operatorname{par} \partial f(\bar{x}) \times \{0\}$. Then there exist $v_1, \dots, v_r \in \partial f(\bar{x})$ and $\lambda_1, \dots, \lambda_{r-1} \in \mathbb{R}$ such that

$$v = \sum_{i=1}^{r-1} \lambda_i(v_i - v_r).$$

Setting $\lambda_r := -\sum_{i=1}^{r-1} \lambda_i$, in view of (13), we find that

$$(v, 0) = \sum_{i=1}^{r} \lambda_i(v_i, -1) \in \operatorname{span} N_{\operatorname{epi} f}(\bar{x}, f(\bar{x})).$$

This concludes the proof. $\qquad\square$

2.2. **Tools from set-valued analysis.** Let $S \subset \mathbb{X}$ and $\bar{x} \in S$. We write $x_k \xrightarrow{S} \bar{x}$ if $\{x_k\} \to \bar{x}$ and $x_k \in S$ for all $k \in \mathbb{N}$. The *regular normal cone* to S at $\bar{x}$ is the (convex) cone defined by

$$\hat{N}_S(\bar{x}) := \left\{ y \in \mathbb{X} \ \middle| \ \limsup_{x \xrightarrow{S} \bar{x}} \frac{\langle y, x - \bar{x} \rangle}{\|x - \bar{x}\|} \leqslant 0 \right\}.$$

It is often expedient to realize that $y \in \hat{N}_S(\bar{x})$ if and only if

$$\forall \varepsilon > 0 \ \exists \delta > 0 \ \forall x \in B_\delta(\bar{x}) \cap S : \ \langle y, x - \bar{x} \rangle \leqslant \varepsilon \|x - \bar{x}\|. \tag{14}$$

The *limiting normal cone* [20, 27] to S at $\bar{x} \in S$ is defined by

$$N_S(\bar{x}) := \left\{ y \in \mathbb{X} \ \middle| \ \exists \{x_k\} \xrightarrow{S} \bar{x}, \ \{y_k \in \hat{N}_S(x_k)\} : \ y_k \to y \right\}, \tag{15}$$

i.e., it is the *outer limit* of the regular normal cone. When S is convex, both the regular and the limiting normal cone reduce to the object in (12). The *tangent cone* to S at $\bar{x}$ is defined by

$$T_S(\bar{x}) := \{w \in \mathbb{X} \mid \exists \{t_k\} \downarrow 0, \{w_k\} \to w : \bar{x} + t_k w_k \in S\}. \tag{16}$$

Let us recall next the limiting and singular subdifferentials of an extended real-valued (possibly nonconvex) function $f : \mathbb{X} \to \mathbb{R} \cup \{\infty\}$ that is lower semicontinuous at some $\bar{x} \in \text{dom} f$; see, e.g., [20, Definition 1.77]. The *limiting subdifferential* of f at $\bar{x}$ is defined by

$$\partial f(\bar{x}) := \{y \in \mathbb{X} |\ (y, -1) \in N_{\text{epi} f}(\bar{x}, f(\bar{x}))\}, \tag{17}$$

which returns the subdifferential of convex analysis defined in Section 2.1 when f is a convex function. Moreover, the *singular subdifferential* of f at $\bar{x}$ is defined by

$$\partial^{\infty} f(\bar{x}) := \{y \in \mathbb{X} |\ (y, 0) \in N_{\text{epi} f}(\bar{x}, f(\bar{x}))\}, \tag{18}$$

which is the horizon cone of $\partial f(\bar{x})$ when f is convex, cf. [27, Proposition 8.12].

An important tool from set-valued differentiation to our study is the *(limiting) coderivative* [20, 27]. Let $F : \mathbb{X} \rightrightarrows \mathbb{Y}$ be a set-valued mapping between two Euclidean spaces. The *graph* of F is defined by

$$\text{gph} F := \{(x, y) \in \mathbb{X} \times \mathbb{Y} \mid y \in F(x)\}.$$

The (limiting) coderivative of F at $\bar{x}$ for $\bar{y} \in F(\bar{x})$ is the set-valued mapping $D^* F(\bar{x} | \bar{y}) : Y \rightrightarrows X$ defined by

$$D^* F(\bar{x} | \bar{y})(w) := \{z \in \mathbb{X} \mid (z, -w) \in N_{\text{gph} F}(\bar{x}, \bar{y})\}. \tag{19}$$

The set-valued mapping F is said to have a *single-valued (and Lipschitz continuous) localization at $\bar{x}$ for $\bar{y}$* (cf. [2, 9, 33, 21, 28]) if there exist open neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that the *localization $s : U \to V$* of F with gph s = gph $F \cap (U \times V)$ is single-valued (and Lipschitz continuous).

We are interested in single-valued locally Lipschitz continuous functions. To make the transition from single-valued localizations to actual single-valuedness for convex-valued mapping, the following result is useful.

Lemma 2.2. *Let $F : \mathbb{X} \rightrightarrows \mathbb{Y}$ be a set-valued mapping, and let $(\bar{x}, \bar{y}) \in \text{gph} F$. Suppose that the set $F(x)$ is convex when x is around $\bar{x}$. Then the following are equivalent:*

(i) *F has a single-valued localization at $\bar{x}$ for $\bar{y}$;*
(ii) *F is single-valued around $\bar{x}$.*

Proof. [(i)$\Rightarrow$ (ii)]: If there exist neighborhoods U of $\bar{x}$ and V of $\bar{y}$ such that the *localization $s : U \to V$* of F with gph s = gph $F \cap (U \times V)$ is single-valued, we may suppose w.l.o.g. that $F(x)$ is convex for any $x \in U$ (otherwise shrink the neighborhood). Pick any $y \in F(x)$. As $F(x)$ is convex and $s(x) \in F(x)$, we find some $\lambda \in (0, 1)$ sufficiently small such that $s(x) + \lambda(y - s(x)) \in F(x) \cap V$, which implies that $s(x) + \lambda(y - s(x)) = s(x)$, i.e., $y = s(x)$. Hence $F(x) = s(x)$ for $x \in U$, which means F is single-valued around $\bar{x}$. The opposite implication is straightforward. $\square$

Corollary 2.3. *Let $h : \mathbb{X} \times \mathbb{P} \to \mathbb{R} \cup \{\infty\}$ be a proper function such that $h(\cdot, p)$ is convex for all $p \in \mathbb{P}$. Define $S : \mathbb{P} \to \mathbb{X}$ by*

$$S(p) := \underset{x \in \mathbb{X}}{\text{argmin}}\ h(x, p).$$

Let $(\bar{p}, \bar{x}) \in \text{gph} S$. Then S has a single-valued (Lipschitz) localization at $\bar{p}$ for $\bar{x}$ if and only if S is single-valued (and locally Lipschitz) around $\bar{x}$.

Proof. This follows immediately from Lemma 2.2, since S is convex-valued. $\square$

2.3. Tools from parametric optimization.
For Euclidean spaces $\mathbb{U}$, $\mathbb{V}$, and $\mathbb{P}$, consider the following parametric optimization problem

$$\text{OPT}(p): \quad \min_{u \in \mathbb{U}} \quad \varphi(u, p) \quad \text{s.t.} \quad G(u, p) \in \Theta, \tag{20}$$

where $\Theta \subset \mathbb{V}$ is a closed and convex set, and $\varphi : \mathbb{U} \times \mathbb{P} \to \mathbb{R}$ and $G : \mathbb{U} \times \mathbb{P} \to \mathbb{V}$ are twice continuously differentiable functions. *Robinson's constraint qualification* is said to hold for $\mathrm{OPT}(\bar{p})$ at some $\bar{u} \in \mathbb{U}$ with $G(\bar{u}, \bar{p}) \in \Theta$ (i.e., $\bar{u}$ is feasible for $\mathrm{OPT}(\bar{p})$) if

$$0 \in \mathrm{int}\, (G(\bar{u}, \bar{p}) + \mathrm{rge}\, \mathcal{J}_u G(\bar{u}, \bar{p}) - \Theta). \tag{21}$$

The *Lagrangian* of problem $\mathrm{OPT}(p)$ is $\mathscr{L} : \mathbb{U} \times \mathbb{P} \times \mathbb{V} \to \mathbb{R}$ given by

$$\mathscr{L}(u, p, \lambda) := \varphi(u, p) + \langle \lambda,\, G(u, p) \rangle \quad \text{for} \quad (u, p, \lambda) \in \mathbb{U} \times \mathbb{P} \times \mathbb{V}. \tag{22}$$

A feasible point $u \in \mathbb{U}$ is called a *stationary* point of $\mathrm{OPT}(p)$ if there exists a *Lagrange multiplier* $\lambda \in \mathbb{V}$ satisfying

$$0 = \nabla_u \mathscr{L}(u, p, \lambda) \quad \text{and} \quad \lambda \in N_\Theta(G(u, p)). \tag{23}$$

We define $\Lambda(u, p) := \{\lambda \in \mathbb{V} \mid 0 = \nabla_u \mathscr{L}(u, p, \lambda),\, \lambda \in N_\Theta(G(u, p))\}$, the set of all Lagrange multipliers at (u, p).

Observe that the system (23) can be written as the *generalized equation* [28]:

$$\mathrm{GE}(p): \quad 0 \in \begin{pmatrix} \nabla_u \varphi(u, p) + \mathcal{J}_u G(u, p)^* \lambda \\ -G(u, p) \end{pmatrix} + \begin{pmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{pmatrix}. \tag{24}$$

Let $H : \mathbb{P} \rightrightarrows \mathbb{U} \times \mathbb{V}$ be its *solution mapping*, i.e., $H(p) = \{(u, \lambda) \mid (u, p, \lambda) \text{ satisfies } (24)\}$. Given $\bar{p} \in \mathbb{P}$ and $(\bar{u}, \bar{\lambda}) \in H(\bar{p})$, it is an immediate question under which conditions H has a single-valued and Lipschitz continuous localization at $\bar{p}$ for $(\bar{u}, \bar{\lambda})$ (in the sense of the definition at the end of the previous section). To answer this question, Robinson [28] came up with an original idea of considering the *linearized generalized equation* of (24) below:

$$\delta \in \begin{pmatrix} 0 \\ -G(\bar{u}, \bar{p}) \end{pmatrix} + \begin{pmatrix} \nabla_{uu}^2 \mathscr{L}(\bar{u}, \bar{p}, \bar{\lambda})(u - \bar{u}) + \mathcal{J}_u G(\bar{u}, \bar{p})^*(\lambda - \bar{\lambda}) \\ -\mathcal{J}_u G(\bar{u}, \bar{p})(u - \bar{u}) \end{pmatrix} + \begin{pmatrix} 0 \\ N_\Theta^{-1}(\lambda) \end{pmatrix} \tag{25}$$

with $\delta \in \mathbb{U} \times \mathbb{V}$. Let us set the solution mapping of this generalized equation by $K : \mathbb{U} \times \mathbb{V} \rightrightarrows \mathbb{U} \times \mathbb{V}$ with $K(\delta) = \{(u, \lambda) \mid (u, \lambda, \delta) \text{ satisfies } (25)\}$ for any $\delta \in \mathbb{U} \times \mathbb{V}$. If K has a single-valued and Lipschitz continuous localization around $\bar{\delta} = 0$ for $(\bar{u}, \bar{\lambda})$, then H also has a single-valued and Lipschitz continuous localization around $\bar{p}$ for $(\bar{u}, \bar{\lambda})$. This is usually referred to as Robinson's implicit function theorem; see the original result by Robinson [28, Theorem 2.1], and its extension [9, Theorem 2B.5] that can be applied directly to our above framework. Lipschitz continuity of the solution mapping of system (25) is known as *Robinson's strong regularity* defined precisely next.

Definition 2.4 (Robinson's strong regularity)**.** *Suppose that $(\bar{u}, \bar{\lambda})$ is a solution of the generalized equation $\mathrm{GE}(\bar{p})$ in (24), i.e., $\bar{u}$ is a stationary point of problem $\mathrm{OPT}(\bar{p})$ in (20) and $\bar{\lambda}$ is a corresponding Lagrange multiplier from (23). We say that the generalized equation $\mathrm{GE}(\bar{p})$ is strongly regular at $(\bar{u}, \bar{\lambda})$ if the solution mapping K of the linearized generalized equation (25) has a single-valued and Lipschitz continuous localization around $0 \in \mathbb{U} \times \mathbb{V}$ for $(\bar{u}, \bar{\lambda}) \in \mathbb{U} \times \mathbb{V}$.*

Robinson's strong regularity was introduced in [28] for generalized equations with applications to nonlinear programming at which $\mathbb{U} = \mathbb{R}^n$, $\mathbb{P} = \mathbb{R}^d$, $\mathbb{V} = \mathbb{R}^{m_1+m_2}$, and $\Theta = \mathbb{R}_-^{m_1} \times \{0_{m_2}\} \subset \mathbb{R}^{m_1+m_2}$. Particularly, Robinson showed in [28, Theorem 4.1] that the well-known *Linear Independence Constraint Qualification* (LICQ) of $\mathcal{J}_u G(\bar{u}, \bar{p})$ and the so-called *Strong Second-Order Sufficient Condition* (SSOSC) that is also proposed by Robinson are sufficient for strong regularity of the corresponding generalized equation (24) at $(\bar{u}, \bar{\lambda})$. If additionally $\bar{u}$ is an optimal solution of problem $\mathrm{OPT}(\bar{p})$ in (20), these two conditions are also necessary for strong regularity; see, e.g., [16, Theorem 8.10] and also [2, 8, 9] for further historic discussions about strong regularity. For semidefinite programming (SDP), Robinson's strong regularity is characterized via different conditions including the so-called *constraint nondegeneracy* and the *SDP-Strong Second-Order Sufficient Condition* by Sun [33]. For more general constrained programming as in (20), it is shown by Mordukhovich, Nghia, and Rockafellar in [21] that Robinson's strong regularity is equivalent to the *Partial Constraint*

Nondegeneracy, which originated from [29] and the *generalized SSOSC* recalled later, provided that the set Θ is $\mathcal{C}^2$-*cone reducible*; see, e.g., [2, Section 3.4.4]. Let us recall these definitions here.

Definition 2.5 (Partial Constraint Nondegeneracy). *Let $\bar{p} \in \mathbb{P}$ and let $\bar{u}$ be feasible for* $\mathrm{OPT}(\bar{p})$, *i.e.,* $G(\bar{u}, \bar{p}) \in \Theta$. *We say that the point $\bar{u}$ is* partially nondegenerate *for G with respect to Θ if*

$$\mathrm{rge}\,\mathcal{J}_u G(\bar{u}, \bar{p}) + \mathrm{lin}\,T_\Theta(G(\bar{u}, \bar{p})) = \mathbb{V}, \tag{26}$$

where $\mathrm{lin}\,T_\Theta(G(\bar{u}, \bar{p}))$ *is the lineality space of* $T_\Theta(G(\bar{u}, \bar{p}))$, *the largest subspace in the tangent cone* $T_\Theta(G(\bar{u}, \bar{p}))$.

Remark 2.6 (Constraint nondegeneracy). *(a) We note that for the case of nonlinear programming when* $\mathbb{U} = \mathbb{R}^n$, $\mathbb{P} = \mathbb{R}^d$, $\mathbb{V} = \mathbb{R}^{m_1+m_2}$, *and* $\Theta = \mathbb{R}^{m_1}_- \times \{0_{m_2}\} \subset \mathbb{R}^{m_1+m_2}$, *and* $G(\bar{x}) \in \Theta$, *the constraint nondegeneracy at $\bar{u}$ holds if and only if the* linear independence constraint qualification (LICQ) *holds at $\bar{u}$.*

(b) A dual formulation of (26) *reads*

$$\ker \mathcal{J}_u G(\bar{u}, \bar{p})^* \bigcap \mathrm{span}\,N_\Theta(G(\bar{u}, \bar{p})) = \{0\}. \tag{27}$$

Recall that a cone $C \subset \mathbb{X}$ is called *pointed* if $C \cap (-C) = \{0\}$.

Definition 2.7 ($\mathcal{C}^2$-cone reducible sets and functions). *The set $\Theta \subset \mathbb{V}$ is said to be $\mathcal{C}^2$-cone reducible at $\bar{v}$ if there exist a neighborhood V of $\bar{v}$, a pointed, closed, convex cone C in some Euclidean space $\mathbb{W}$, and a $\mathcal{C}^2$-smooth mapping $h : V \to \mathbb{W}$ such that $h(\bar{v}) = 0$, $\mathcal{J}h(\bar{v}) \in \mathcal{L}(\mathbb{V}, \mathbb{W})$ is surjective, and that $\Theta \cap V = \{v \in V \mid h(v) \in C\}$.*

A closed, proper, convex function $f : \mathbb{U} \to \mathbb{R} \cup \{+\infty\}$ is called $\mathcal{C}^2$-cone reducible at $\bar{u} \in \mathrm{dom}\,f$ if its epigraph $\mathrm{epi}\,f$ is $\mathcal{C}^2$-cone reducible at $(\bar{u}, f(\bar{u}))$. We call the function f to be $\mathcal{C}^2$-cone reducible, if it is $\mathcal{C}^2$-cone reducible at any point on its domain.

We point out that any set Θ is $\mathcal{C}^2$-cone reducible at any $\bar{v} \in \mathrm{int}\,\Theta$, as we may choose a neighborhood $V \subset \Theta$ of $\bar{v}$, $\mathbb{W} = \{0\}$, $C = \{0\}$, and $h(v) \equiv 0$ in the above definition. Many important convex sets in optimization are $\mathcal{C}^2$-cone reducible including (convex) polyhedra [2, Example 3.139], the positive semidefinite cone [2, Example 3.140], or the second-order (or Lorentz) cone [1, Lemma 15]. The Cartesian product of many $\mathcal{C}^2$-cone reducible convex sets is also $\mathcal{C}^2$-cone reducible; see [31, Proposition 3.1]. Moreover, it is shown in [10, Proposition 10] that many *spectral functions* including the nuclear norm and the spectral norm are also $\mathcal{C}^2$-cone reducible.

The following result [21, Theorem 5.6] provides a full characterization of Robinson's strong regularity for $\mathcal{C}^2$-cone reducible programming.

Theorem 2.8 (Characterization of Robinson's strong regularity). *Let $\bar{u} \in \mathbb{U}$ be a stationary point of problem $\mathrm{OPT}(\bar{p})$ in (20) with $\bar{\lambda} \in \Lambda(\bar{u}, \bar{p})$. Suppose that the set Θ is $\mathcal{C}^2$-cone reducible at $G(\bar{u}, \bar{p})$ and that Robinson's constraint qualification (21) holds at $\bar{u}$. Then the following are equivalent:*

  (i) *The generalized equation (24) is strongly regular at $(\bar{u}, \bar{\lambda})$ and $\bar{u}$ is a local minimizer of problem $\mathrm{OPT}(\bar{p})$;*

  (ii) *$\bar{u}$ is partially nondegenerate for G with respect to Θ and the following so-called* generalized strong second-order sufficient condition (GSSOSC) *holds*

$$\langle w, \nabla^2_{uu}\mathscr{L}(\bar{u}, \bar{p}, \bar{\lambda})w \rangle + \inf\left\{ \langle z, \mathcal{J}_u G(\bar{u}, \bar{p})w \rangle \mid z \in D^* N_\Theta(G(\bar{u}, \bar{p}) \mid \bar{\lambda})\big(\mathcal{J}_u G(\bar{u}, \bar{p})w\big) \right\} > 0 \tag{28}$$

  for any $w \neq 0$.

In the framework of nonlinear programming, the GSSOSC is exactly the classical SSOSC of Robinson [28]. For the case $\Theta = \mathbb{S}^n_+$, the cone of of positive semi-definite matrices, it also recovers the SDP-SSOSC [33]; see [21] for further details and discussions. We point out that verifying GSSOSC (28) is not trivial due to the lack of explicit calculus for the coderivative of the normal cone mapping. In Section 3, however, we will show that GSSOSC holds automatically for the problems in the center of our study.

3. LIPSCHITZ STABILITY OF LEAST-SQUARES PROBLEMS WITH C^2-CONE REDUCIBLE DUAL REGULARIZES

In this section, we consider the following parametric optimization problem

$$P(A, b, \mu): \quad \min_{x \in \mathbb{X}} \quad \frac{1}{2\mu}\|Ax - b\|^2 + g(x), \tag{29}$$

where $\mathbb{X}, \mathbb{Y}$ are two Euclidean spaces, $A \in \mathcal{L}(\mathbb{X}, \mathbb{Y})$, $b \in \mathbb{Y}$, $\mu > 0$ are treated as parameters and $g : \mathbb{X} \to \mathbb{R} \cup \{+\infty\}$ is a closed, proper, convex function. Define the solution mapping $S : \mathcal{L}(\mathbb{X}, \mathbb{Y}) \times \mathbb{Y} \times \mathbb{R}_{++} \to \mathbb{X}$ of problem (29) by

$$S(A, b, \mu) := \operatorname*{argmin}_{x \in \mathbb{X}} \left\{ \frac{1}{2\mu}\|Ax - b\|^2 + g(x) \right\}. \tag{30}$$

Our main goal in this section is to study sufficient and necessary conditions for this solution mapping to be (locally) Lipschitz around some fixed triple $(\bar{A}, \bar{b}, \bar{\mu}) \in \mathcal{L}(\mathbb{X}, \mathbb{Y}) \times \mathbb{Y} \times \mathbb{R}_{++}$. Given $v \in \mathbb{X}$, the *tilt-perturbed* problem corresponding to $P(A, b, \mu)$ reads

$$Q(A, b, \mu, v): \quad \min_{x \in \mathbb{X}} \quad \frac{1}{2\mu}\|Ax - b\|^2 + g(x) - \langle v,\, x \rangle. \tag{31}$$

Its solution mapping $\widehat{S} : \mathcal{L}(\mathbb{X}, \mathbb{Y}) \times \mathbb{Y} \times \mathbb{R}_{++} \times \mathbb{X} \to \mathbb{X}$ is defined by

$$\widehat{S}(A, b, \mu, v) = \operatorname*{argmin}_{x \in \mathbb{X}} \left\{ \frac{1}{2\mu}\|Ax - b\|^2 + g(x) - \langle v,\, x \rangle \right\}. \tag{32}$$

Obviously, $Q(A, b, \mu, 0)$ and $P(A, b, \mu)$ are the same problem, and thus $\widehat{S}(A, b, \mu, 0) = S(A, b, \mu)$. Adding the linear perturbation $\langle v,\, x \rangle$ in the problem will be useful for Section 4 where we leverage our analysis here for more general problems.

Proposition 3.1 (Primal-dual optimality). *The (Fenchel-Rockafellar) dual of problem* (31) *reads*

$$D(A, b, \mu, v): \quad \min_{y \in \mathbb{Y}} \quad \frac{\mu}{2}\|y\|^2 - \langle b,\, y \rangle + g^*(A^*y + v). \tag{33}$$

For $(A, b, \mu, v) \in \mathcal{L}(\mathbb{X}, \mathbb{Y}) \times \mathbb{Y} \times \mathbb{R}_{++} \times \mathbb{X}$, the following are equivalent:

   (i) *x solves $Q(A, b, \mu, v)$ and y solves $D(A, b, \mu, v)$;*
   (ii) *$A^*y + v \in \partial g(x)$ and $-\mu y = Ax - b$;*
   (iii) *$x \in \partial g^*(A^*y + v)$ and $-\mu y = Ax - b$.*

Proof. We apply the duality scheme from [27, Example 11.41] with $h = \frac{1}{\mu}\|\cdot\|^2$, $k = g$ and $c = -v$. The only thing we have to ensure is that primal and dual optimal value are equal which is guaranteed as $\operatorname{dom} h = \mathbb{Y}$, cf. [27, Theorem 11.39]. $\qquad\square$

Note that an equivalent formulation of the dual problem (33) reads

$$D'(A, b, \mu, v): \quad \min_{(y,t) \in \mathbb{Y} \times \mathbb{R}} \quad \frac{\mu}{2}\|y\|^2 - \langle b,\, y \rangle + t \quad \text{s.t.} \quad (A^*y + v, t) \in \operatorname{epi} g^*. \tag{34}$$

Although the parameter μ is chosen to be positive for problem (29), problem (34) is well-defined for any $\mu \in \mathbb{R}$. We will apply the theory laid out for the parametric problem $\mathrm{OPT}(p)$ in (20) to problem $D'(A, b, \mu, v)$ above with the spaces $\mathbb{U} = \mathbb{Y} \times \mathbb{R}$, $\mathbb{V} = \mathbb{X} \times \mathbb{R}$, $\mathbb{P} = \mathcal{L}(\mathbb{X}, \mathbb{Y}) \times \mathbb{Y} \times \mathbb{R} \times \mathbb{X}$, the parameter $p = (A, b, \mu, v) \in \mathbb{P}$, the variable $u = (y, t) \in \mathbb{U}$, the convex set $\Theta := \operatorname{epi} g^* \subset \mathbb{V}$, and the functions

$$\varphi(u, p) := \frac{\mu}{2}\|y\|^2 - \langle b,\, y \rangle + t \quad \text{and} \quad G(u, p) := (A^*y + v, t). \tag{35}$$

To this end, we first give an appropriate expression for the partial constraint nondegeneracy of $D'(A, b, \mu, 0)$.

Lemma 3.2 (Partial constraint nondegeneracy for dual problem)**.** *Let $\bar{u} = (\bar{y}, \bar{t})$ be a solution of $D'(\bar{p})$ with fixed $\bar{p} = (\bar{A}, \bar{b}, \bar{\mu}, 0) \in \mathbb{P}$ and $\bar{\mu} > 0$. Then the following are equivalent:*

   (i) *The partial constraint nondegeneracy holds at $(\bar{y}, \bar{t})$ for $D'(\bar{A}, \bar{b}, \bar{\mu}, 0)$;*
   (ii) $\ker \bar{A} \cap \operatorname{par} \partial g^*(\bar{A}^* \bar{y}) = \{0\}$.

Proof. Set $\bar{z} := \bar{A}^* \bar{y}$. In view of Remark 2.6(b) and with the functions in (35), we find that constraint nondegeneracy holds for $D'(\bar{p})$ at $(\bar{y}, \bar{t})$ if and only if

$$\{0\} = \ker \bar{A} \times \{0\} \bigcap \operatorname{span} N_{\operatorname{epi} g^*}(\bar{z}, \bar{t}). \tag{36}$$

Now, note that, by optimality, we necessarily have $\bar{t} = g^*(\bar{z})$. From Proposition 2.1, we know that

$$\mathbb{X} \times \{0\} \bigcap \operatorname{span} N_{\operatorname{epi} g^*}(\bar{z}, \bar{t}) = \operatorname{par} \partial g^*(\bar{z}) \times \{0\}.$$

This yields the desired equivalence. $\qquad\square$

The Lagrangian of $D'(A, b, \mu, v)$ in the sense of (22) reads

$$\mathscr{L}(u, p, \lambda) = \frac{\mu}{2}\|y\|^2 - \langle b,\, y \rangle + t + \langle \lambda_1,\, A^* y + v \rangle + \lambda_2 t \tag{37}$$

with $u = (y, t) \in \mathbb{U} = \mathbb{Y} \times \mathbb{R}$, $p = (A, b, \mu, v) \in \mathbb{P}$, and $\lambda = (\lambda_1, \lambda_2) \in \mathbb{V} = \mathbb{X} \times \mathbb{R}$. Consequently, stationarity conditions for $D'(A, b, \mu, v)$ in the sense of (23) read

$$0 = \mu y - b + A\lambda_1, \quad 0 = 1 + \lambda_2, \quad (\lambda_1, \lambda_2) \in N_{\operatorname{epi} g^*}(A^* y + v, t), \tag{38}$$

which certainly tells us that $\lambda_2 = -1$. Combining (38) and Proposition 3.1 yields the following result.

Proposition 3.3. *For $(A, b, \mu, v) \in \mathcal{L}(\mathbb{X}, \mathbb{Y}) \times \mathbb{Y} \times \mathbb{R}_{++} \times \mathbb{X}$, the following are equivalent:*

   (i) *x solves $Q(A, b, \mu, v)$ and y solves $D(A, b, \mu, v)$;*
   (ii) *$(y, g^*(A^* y + v))$ solves $D'(A, b, \mu, v)$ and $(x, -1)$ is a corresponding Lagrange multiplier of problem $D'(A, b, \mu, v)$.*

Proof. The proof follows from combining Proposition 3.1 and the stationarity conditions in (38) for $D'(A, b, \mu, v)$. Here, it is worth noting that $(x, -1) \in N_{\operatorname{epi} g^*}(A^* y + v, g^*(A^* y + v))$ if and only if $x \in \partial g^*(A^* y + v)$ by (13). $\qquad\square$

To prove Lipschitz stability of the solution map $\widehat{S}$ around $\bar{p} = (\bar{A}, \bar{b}, \bar{\mu}, \bar{v})$, in view of Proposition 3.3, it suffices to show that the solution mapping of the stationarity system (38) is locally Lipschitz (in the Lagrange multipliers). By the discussion before Definition 2.4, the latter occurs when Robinson's strong regularity holds at $(\bar{u}, \bar{\lambda})$ for the dual problem $D'(\bar{p})$. By Theorem 2.8, we need that both the partial constraint nondegeneracy (26) and the generalized strong second-order sufficient condition (28) are satisfied for the system (38). First we show that the generalized SSOSC (38) is automatic for the dual problem $D'(\bar{p})$.

Proposition 3.4 (Validity of Generalized SSOSC)**.** *Let $\bar{u} = (\bar{y}, \bar{t})$ be a solution of $D'(\bar{p})$ with $\bar{p} = (\bar{A}, \bar{b}, \bar{\mu}, 0)$ and $\bar{\mu} > 0$. Suppose that $\bar{\lambda} = (\bar{\lambda}_1, -1)$ is the corresponding Lagrange multiplier at $(\bar{y}, \bar{t})$ satisfying (38). Then the Generalized SSOSC (28) for the dual problem $D'(\bar{p})$ holds in the sense that*

$$\langle w, \nabla_{uu}^2 \mathscr{L}(\bar{u}, \bar{p}, \bar{\lambda}) w \rangle + \inf \left\{ \langle z,\, \mathcal{J}_u G(\bar{u}, \bar{p}) w \rangle \mid z \in D^* N_\Theta(G(\bar{u}, \bar{p}), \bar{\lambda})(\mathcal{J}_u G(\bar{u}, \bar{p}) w) \right\} > 0 \tag{39}$$

for any $w \in \mathbb{U} \backslash \{0\}$ with all the notations in (35) and (37).

Proof. Note that, by (37), we have

$$\nabla_{uu}^2 \mathscr{L}(\bar{u}, \bar{p}, \bar{\lambda}) w = (\bar{\mu} w_1, 0) \quad \forall w = (w_1, w_2) \in \mathbb{Y} \times \mathbb{R}. \tag{40}$$

In particular, $\nabla_{uu}^2 \mathscr{L}(\bar{u}, \bar{p}, \bar{\lambda})$ is positive semidefinite. Moreover, since the normal cone mapping N_Θ is *maximally monotone*, we have

$$\inf_{z \in \mathbb{X} \times \mathbb{R}} \left\{ \langle z, \mathcal{J}_u G(\bar{u}, \bar{p}) w \rangle \,|\, z \in D^* N_\Theta(G(\bar{u}, \bar{p})| \bar{\lambda})(\mathcal{J}_u G(\bar{u}, \bar{p}) w) \right\} \geqslant 0, \tag{41}$$

see, e.g., [25, Theorem 3.1]. Therefore, the left-hand side of (39) is nonnegative for any $w \in \mathbb{U}$. It hence suffices to show that, if it is equal to zero, then $w = 0$. To this end, let $w \in \mathbb{U}$ satisfy

$$\langle w, \nabla_{uu}^2 \mathscr{L}(\bar{u}, \bar{p}, \bar{\lambda}) w \rangle + \inf_{z \in \mathbb{X} \times \mathbb{R}} \left\{ \langle z, \mathcal{J}_u G(\bar{u}, \bar{p}) w \rangle \,|\, z \in D^* N_\Theta(G(\bar{u}, \bar{p})| \bar{\lambda})(\mathcal{J}_u G(\bar{u}, \bar{p}) w) \right\} = 0. \tag{42}$$

Applying (41) and (40) to (42) yields $w_1 = 0$. And consequently, we find that

$$\mathcal{J}_u G(\bar{u}, \bar{p}) w = (A^* w_1, w_2) = (0, w_2). \tag{43}$$

In particular, (42) reduces to

$$\inf_{(z_1, z_2) \in \mathbb{X} \times \mathbb{R}} \left\{ z_2 w_2 \,|\, (z_1, z_2) \in D^* N_\Theta(G(\bar{u}, \bar{p})| \bar{\lambda})(0, w_2) \right\} = 0.$$

It remains to show that $w_2 = 0$. To this end, pick $(z_1, z_2) \in D^* N_\Theta(G(\bar{u}, \bar{p})| \bar{\lambda})(0, w_2)$, i.e.,

$$((z_1, z_2), (0, -w_2)) \in N_{\mathrm{gph}\, N_\Theta}(G(\bar{u}, \bar{p}), \bar{\lambda}).$$

By the definition of the normal cone (15), we find sequences $(v^k, t_k) \to G(\bar{u}, \bar{p}) = (\bar{A}^* \bar{y}, \bar{t})$, $(\lambda_1^k, \lambda_2^k) \to \bar{\lambda} = (\bar{\lambda}_1, -1)$, $(z_1^k, z_2^k) \to (z_1, z_2)$, and $(w_1^k, w_2^k) \to (0, w_2)$ such that $g^*(v^k) \leqslant t_k$, $(\lambda_1^k, \lambda_2^k) \in N_\Theta(v^k, t_k)$ and

$$\left((z_1^k, z_2^k), (w_1^k, -w_2^k) \right) \in \hat{N}_{\mathrm{gph}\, N_\Theta} \left((v^k, t_k), (\lambda_1^k, \lambda_2^k) \right).$$

By the expression of regular normal vectors in (14), for all $\epsilon > 0$ there exists $\delta > 0$ such that for any $((v, t), (\lambda_1, \lambda_2)) \in \mathrm{gph}\, N_\Theta \cap \mathbb{B}_\delta((v^k, t_k), (\lambda_1^k, \lambda_2^k))$:

$$\left\langle z_1^k, v - v^k \right\rangle + z_2^k(t - t_k) + \left\langle w_1^k, \lambda_1 - \lambda_1^k \right\rangle - w_2^k(\lambda_2 - \lambda_2^k) \leqslant \epsilon \|(v, t, \lambda_1, \lambda_2) - (v^k, t_k, \lambda_1^k, \lambda_2^k)\|. \tag{44}$$

As $(\lambda_1^k, \lambda_2^k) \in N_{\mathrm{epi}\, g*}(v^k, t_k)$, we get from (12) that

$$\left\langle \lambda_1^k, v - v^k \right\rangle + \lambda_2^k(t - t_k) \leqslant 0 \quad \text{for all} \quad (v, t) \in \mathrm{epi}\, g^*.$$

By choosing $(v, t) = (v^k, g^*(v^k))$ in the above inequality, it follows that $\lambda_2^k(g^*(v^k) - t_k) \leqslant 0$. As λ_2^k is close to -1 and $g^*(v^k) \leqslant t_k$, we have $t_k = g^*(v^k)$.

Let us choose (λ_1, λ_2) satisfying $\dfrac{\lambda_1}{\lambda_2} = \dfrac{\lambda_1^k}{\lambda_2^k}$ and such that $(\lambda_1, \lambda_2) \in \mathbb{B}_\delta(\lambda_1^k, \lambda_2^k)$. It follows from formula (13) (which is applicable as $t_k = g^*(v^k)$) that

$$\frac{\lambda_1}{-\lambda_2} = \frac{\lambda_1^k}{-\lambda_2^k} \in \partial g^*(v^k).$$

This implies that $(\lambda_1, \lambda_2) \in N_{\mathrm{epi}\, g*}(v^k, t_k)$, i.e., $((v^k, t_k), (\lambda_1, \lambda_2)) \in \mathrm{gph}\, N_\Theta \cap \mathbb{B}_\delta((v^k, t_k), (\lambda_1^k, \lambda_2^k))$. We can thus insert $\lambda_1 = \lambda_2 \dfrac{\lambda_1^k}{\lambda_2^k}$ and $v = v^k, t = t_k$ in (44) to obtain

$$\frac{\lambda_2 - \lambda_2^k}{\lambda_2^k} \left\langle w_1^k, \lambda_1^k \right\rangle - w_2^k(\lambda_2 - \lambda_2^k) \leqslant \epsilon \left[\|\lambda_1^k\| \frac{|\lambda_2 - \lambda_2^k|}{|\lambda_2^k|} + |\lambda_2 - \lambda_2^k| \right],$$

which implies that

$$-w_2^k(\lambda_2 - \lambda_2^k) \leqslant \epsilon \left[\|\lambda_1^k\| \frac{|\lambda_2 - \lambda_2^k|}{|\lambda_2^k|} + |\lambda_2 - \lambda_2^k| \right] + \frac{|\lambda_2 - \lambda_2^k|}{|\lambda_2^k|} \cdot \left| \left\langle w_1^k, \lambda_1^k \right\rangle \right|.$$

As λ_2 can be chosen around λ_2^k, the above inequality implies

$$|w_2^k| \leqslant \epsilon \left[\frac{\|\lambda_1^k\|}{|\lambda_2^k|} + 1 \right] + \frac{|\langle w_1^k, \lambda_1^k \rangle|}{|\lambda_2^k|}.$$

Since $(\lambda_1^k, \lambda_2^k) \to (\bar{\lambda}_1, -1)$ and $(w_1^k, w_2^k) \to (0, w_2)$, we get from the above inequality that

$$|w_2| \leqslant \epsilon(\|\bar{\lambda}_1\| + 1) \quad \text{for any} \quad \epsilon > 0.$$

This tells us that $w_2 = 0$, which verifies the claim at the beginning. $\qquad\square$

Combining Lemma 3.2, Proposition 3.3, and Proposition 3.4 with Theorem 2.8, we obtain the sufficient condition for Lipschitz stability of the solution mapping of parametric problems $P(A, b, \mu)$ in (29) and $Q(A, b, \mu, v)$ in (31) below.

Theorem 3.5. *Let $\bar{x}$ be a solution of $P(\bar{A}, \bar{b}, \bar{\mu})$ for $(\bar{A}, \bar{b}, \bar{\mu}) \in \mathcal{L}(\mathbb{X}, \mathbb{Y}) \times \mathbb{Y} \times \mathbb{R}_{++}$ and suppose that g^* is a $\mathcal{C}^2$-cone reducible function at $\bar{z} = \frac{1}{\bar{\mu}} \bar{A}^*(\bar{b} - \bar{A}\bar{x})$. Assume further that*

$$\ker \bar{A} \cap \operatorname{par} \partial g^*(\bar{z}) = \{0\}. \tag{45}$$

Then, the solution mapping $\hat{S}$ defined in (32) is single-valued and Lipschitz continuous around $(\bar{A}, \bar{b}, \bar{\mu}, 0)$ with $\hat{S}(\bar{A}, \bar{b}, \bar{\mu}, 0) = \bar{x}$. Consequently, the solution mapping S of problem (29) is also single-valued and Lipschitz continuous around $(\bar{A}, \bar{b}, \bar{\mu})$ with $S(\bar{A}, \bar{b}, \bar{\mu}) = \bar{x}$.

Proof. Let $\bar{y}$ be the unique solution of the dual problem $D(\bar{A}, \bar{b}, \bar{\mu}, 0)$ in (33). Note from Proposition 3.1 that $\bar{z} = A^* \bar{y}$. By Lemma 3.2, the partial constraint nondegeneracy holds at $(\bar{y}, g^*(\bar{z}))$ for the dual problem $D'(\bar{A}, \bar{b}, \bar{\mu}, 0)$ in (34). By Proposition 3.4, the generalized SSOSC for the dual problem $D'(\bar{A}, \bar{b}, \bar{\mu}, 0)$ is also satisfied. Moreover, since g^* is $\mathcal{C}^2$-cone reducible at $\bar{z}$, its epigraph epi g^* is $\mathcal{C}^2$-cone reducible at $G(\bar{u}, \bar{p})$ from (35) with $\bar{u} = (\bar{y}, g^*(\bar{z}))$ and $\bar{p} = (\bar{A}, \bar{b}, \bar{\mu}, 0)$. We are in the position of applying Theorem 2.8 to the dual problem (34). By the discussion after Definition 2.4, the Lagrange multiplier mapping Λ of $D'(A, b, \mu, v)$ must have a single-valued and Lipschitz continuous localization around $(\bar{A}, \bar{b}, \bar{\mu}, 0)$ for $\bar{x}$. By Proposition 3.3, $\Lambda(A, b, \mu, v) = \hat{S}(A, b, \mu, v) \times \{1\}$. Thus the solution mapping $\hat{S}$ has a single-valued and Lipschitz continuous localization around $(\bar{A}, \bar{b}, \bar{\mu}, 0)$ for $\bar{x}$. The desired statements now follow from Corollary 2.3 and the fact that $S(A, b, \mu) = \hat{S}(A, b, \mu, 0)$. $\qquad\square$

Remark 3.6 (The class of $\mathcal{C}^2$-cone reducible conjugates). Our main assumption on the regularizer g in Theorem 3.5 to have a conjugate g^* that is $\mathcal{C}^2$-cone reducible at $\bar{z}$ appears to be restrictive at first glance. However, it turns out that many important convex regularizers g satisfy this condition. We furnish evidence for this by the following list of (important) examples:

(a) *(Convex piecewise linear functions)* The proper convex function $g : \mathbb{X} \to \mathbb{R} \cup \{\infty\}$ is called convex *piecewise linear* on its domain if dom g can be represented as the union of finitely many polyhedral sets, relative to each of which $g(x)$ is given by an expression $\langle a, x \rangle + b$ for some $a \in \mathbb{X}$ and $b \in \mathbb{R}$; see, e.g., [27, Definition 2.47]. The Fenchel conjugate of g is also a piecewise linear function according to [27, Theorem 11.14]. Thus, the epigraph of g^* is a polyhedra by [27, Theorem 2.49]. It follows from [2, Example 3.139] that epi g^* is $\mathcal{C}^2$-cone reducible. This tells us that the conjugate of any proper convex piecewise linear function is $\mathcal{C}^2$-cone reducible.

(b) *(Support functions of $\mathcal{C}^2$-cone reducible set)* Let g be the *support function* of a closed, convex, and $\mathcal{C}^2$-cone reducible set $C \subset \mathbb{X}$, i.e.,

$$g(x) = \sigma_C(x) := \sup \{\langle v, x \rangle \mid v \in C\}. \tag{46}$$

Note that $g^* = \delta_C$, the indicator function of C. Hence epi $g^* = C \times \mathbb{R}_+$. Obviously, $\mathbb{R}_+$ is polyhedral, hence $\mathcal{C}^2$-cone reducible [2, Example 3.139]. Thus, the Cartesian product $C \times \mathbb{R}_+$ is $\mathcal{C}^2$-cone reducible. It follows that g^* is $\mathcal{C}^2$-cone reducible.

This class of support functions is broad. It includes norm functions that are widely used as regularizers, see the two following examples.

(c) *(ℓ_1/ℓ_2 norm)* Let g be the ℓ_1/ℓ_2 norm in $\mathbb{R}^n$ given by

$$g(x) = \|x\|_{1,2} := \sum_{J \in \mathcal{G}} \|x_J\| \quad \text{for any} \quad x = (x_J)_{J \in \mathcal{G}} \in \mathbb{R}^n, \tag{47}$$

where $\mathcal{G}$ is a partition of $\{1, 2, \ldots, n\}$ with m different groups $J \in \mathcal{G}$ and $\|x_J\|$ is the Euclidean norm at $x_J \in \mathbb{R}^{|J|}$ for any $J \in \mathcal{G}$. We note that this covers the ℓ_1 norm by the partition $\{\{i\} \mid i = 1, \ldots, n\}$.

Now, define the set $C \subset \mathbb{R}^n$ by

$$C := \prod_{J \in \mathcal{G}} \mathbb{B}_J,$$

where $\mathbb{B}_J$ is the (Euclidean) unit ball in $\mathbb{R}^{|J|}$. Then it is easy to see that $g = \sigma_C$. Note that $\mathbb{B}_J = \{v \in \mathbb{R}^{|J|} \mid h(v) := 1 - \|v\|^2 \in \mathbb{R}_+\}$ is $\mathcal{C}^2$-cone reducible at any v with $\|v\| = 1$, as $h(v) = 0$ and $\mathcal{J}h(v) : \mathbb{R}^{|J|} \to \mathbb{R}, x \mapsto -2v^T x$ is surjective. As discussed after Definition 2.7, $\mathbb{B}_J$ is also $\mathcal{C}^2$-cone reducible at any $\bar{v} \in \text{int } \mathbb{B}_J$. Thus, it is $\mathcal{C}^2$-cone reducible at any $\bar{v} \in \mathbb{B}_J$. It follows that the Cartesian product C defined above is $\mathcal{C}^2$-cone reducible. Consequently, $g^* = \delta_C$ is $\mathcal{C}^2$-cone reducible.

(d) *(Nuclear norm)* Another important regularizer is the *nuclear norm* $\|X\|_*$, the sum of all *singular values* of $X \in \mathbb{R}^{m \times n}$ $(m \leqslant n)$. Its Fenchel conjugate is also $\mathcal{C}^2$-cone reducible; see, e.g., [10, Proposition 3.2 and Remark 3.4]. This fact can be explained directly and differently as follows. The nuclear norm can be written as the support function $\|X\|_* = \sigma_{\mathbb{B}}(X)$, where $\mathbb{B} = \{Z \in \mathbb{R}^{m \times n} \mid \|Z\|_2 \leqslant 1\}$ is the *spectral unit ball* in $\mathbb{R}^{m \times n}$ with $\|\cdot\|_2$ being the spectral norm. We show next that $\mathbb{B}$ is $\mathcal{C}^2$-cone reducible at any $\bar{Z}$ on the boundary of $\mathbb{B}$, i.e., $\|\bar{Z}\|_2 = 1$. Note that the spectral unit ball is represented by

$$\mathbb{B} = \left\{ Z \in \mathbb{R}^{m \times n} \mid G(Z) := \mathbb{I}_n - Z^T Z \in \mathbb{S}^n_+ \right\}, \tag{48}$$

where $\mathbb{I}_n$ is the $n \times n$ identity matrix and $\mathbb{S}^n_+$ is the set of all positive semidefinite $n \times n$ matrices. As the positive semidefinite cone $\mathbb{S}^n_+$ is $\mathcal{C}^2$-cone reducible [2, Example 3.140], it suffices to show that $\bar{Z}$ is nondegenerate for G with respect to $\mathbb{S}^n_+$ according to [31, Proposition 3.2], which means

$$\text{rge } \mathcal{J}G(\bar{Z}) + \text{lin } T_{\mathbb{S}^n_+}(G(\bar{Z})) = \mathbb{S}^n, \tag{49}$$

where $\mathbb{S}^n$ is the space of all $n \times n$ symmetric matrices. Suppose the Singular Value Decomposition (SVD) of $\bar{Z}$ is $\bar{Z} = \bar{U} \Sigma \bar{V}^T$, where $\bar{U} \in \mathbb{R}^{m \times m}$ and $\bar{V} \in \mathbb{R}^{n \times n}$ are orthogonal matrices and $\Sigma = (\text{Diag}(\sigma_1, \ldots, \sigma_m) \quad 0)$ has two matrix blocks: the first block of $m \times m$ is the diagonal matrix containing all singular values $\sigma_1 \geqslant \sigma_2 \ldots \geqslant \sigma_m$ of $\bar{Z}$ and the second block is just zero $m \times (n - m)$ matrix. Note that

$$\mathcal{J}G(\bar{Z})X = -\bar{Z}^T X - X^T \bar{Z} = -\bar{V} \Sigma^T \bar{U}^T X - X^T \bar{U} \Sigma \bar{V}^T \quad \text{for any} \quad X \in \mathbb{R}^{m \times n}. \tag{50}$$

Since $\|\bar{Z}\|_2 = 1$, we have $\sigma_1 = 1$. Let p be the multiplicity of σ_1 in $\{\sigma_1, \ldots, \sigma_m\}$, i.e., $1 = \sigma_p > \sigma_{p+1}$ (with $\sigma_{m+1} := 0$). Choose $X = \bar{U} \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \bar{V}^T$ for any $A \in \mathbb{S}^p$, we derive from (50) that

$$\mathcal{J}G(\bar{Z})X = -\bar{V} \Sigma^T \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \bar{V}^T - \bar{V} \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \Sigma \bar{V}^T = -2\bar{V} \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \bar{V}^T.$$

It follows that

$$\mathrm{rge}\, \mathcal{J}G(\bar{Z}) \supset \left\{ \bar{V} \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \bar{V}^T \,|\, A \in \mathbb{S}^p \right\}. \tag{51}$$

Moreover, observe further that

$$
\begin{aligned}
T_{\mathbb{S}_+^n}(G(\bar{Z})) &= T_{\mathbb{S}_+^n}(\bar{V}(\mathbb{I}_n - \Sigma^T\Sigma)\bar{V}^T) \\
&= T_{\mathbb{S}_+^n}\left(\bar{V} \begin{pmatrix} 0_{p\times p} & 0 & 0 \\ 0 & \mathrm{Diag}\,(1-\sigma_{p+1}^2, \ldots, 1-\sigma_m^2) & 0 \\ 0 & 0 & \mathbb{I}_{n-m} \end{pmatrix} \bar{V}^T \right).
\end{aligned}
$$

By the well-known formula of tangent cone of the positive semidefinite cone [2, Example 2.65], we have

$$T_{\mathbb{S}_+^n}(G(\bar{Z})) = \left\{ \bar{V} \begin{pmatrix} A & B \\ B^T & C \end{pmatrix} \bar{V}^T \,|\, A \in \mathbb{S}_+^p, B \in \mathbb{R}^{p\times(n-p)}, C \in \mathbb{S}^{n-p} \right\}.$$

It follows that

$$\mathrm{lin}\, T_{\mathbb{S}_+^n}(G(\bar{Z})) = \left\{ \bar{V} \begin{pmatrix} 0 & B \\ B^T & C \end{pmatrix} \bar{V}^T \,|\, B \in \mathbb{R}^{p\times(n-p)}, C \in \mathbb{S}^{n-p} \right\},$$

which together with (51) verifies (49). Thus, $\bar{Z}$ is nondegenerate for G with respect to $\mathbb{S}_+^n$. The preimage $\mathbb{B} = G^{-1}(\mathbb{S}_+^n)$ is $\mathcal{C}^2$-cone reducible at $\bar{Z}$ due to [31, Proposition 3.2]. As $\mathbb{B}$ is always $\mathcal{C}^2$-cone reducible at point in is interior, it is $\mathcal{C}^2$-cone reducible at any $\bar{Z} \in \mathbb{B}$. This explains the fact that the Fenchel conjugate of the nuclear norm is $\mathcal{C}^2$-cone reducible. There are many more spectral functions whose Fenchel conjugates are also $\mathcal{C}^2$-cone reducible [10, Proposition 3.2].

$\triangle$

We now show that qualification condition (45) is necessary for the solution map $S(\bar{A}, \cdot, \bar{\mu})$ in (30) to be single-valued.

Theorem 3.7 (Necessity of single-valuedness of solution mapping $S(\bar{A}, \cdot, \bar{\mu})$). *Suppose that the solution mapping $S(\bar{A}, b, \bar{\mu})$ in (30) is single-valued around $\bar{b}$ with $\bar{x} = S(\bar{A}, \bar{b}, \bar{\mu})$. Then, with $\bar{z} = \frac{1}{\bar{\mu}}\bar{A}^*(\bar{b} - \bar{A}\bar{x})$, condition (45) is satisfied.*

Proof. Suppose that the solution mapping $S(\bar{A}, b, \bar{\mu})$ in (30) is single-valued around $\bar{b}$ with $\bar{x} = S(\bar{A}, \bar{b}, \bar{\mu})$. By first-order optimality conditions, we have $\bar{z} \in \partial g(\bar{x})$, i.e., $\bar{x} \in \partial g^*(\bar{z})$. Since $\partial g^*(\bar{z})$ is a closed convex set, we have

$$\mathrm{cl}\,[\mathrm{ri}\, \partial g^*(\bar{z})] = \partial g^*(\bar{z}), \tag{52}$$

see, e.g., [26, Theorem 6.3]. It follows that there exists a sequence $\{x_k \in \mathrm{ri}\, \partial g^*(\bar{z})\} \to \bar{x}$. Therefore, we find that

$$b_k := \bar{b} + \bar{A}(x_k - \bar{x}) \to \bar{b}.$$

Observe that

$$\frac{1}{\bar{\mu}}\bar{A}^*(b_k - \bar{A}x_k) = \frac{1}{\bar{\mu}}\bar{A}^*(\bar{b} - \bar{A}\bar{x}) = \bar{z} \in \partial g(x_k).$$

By first-order optimality, we thus infer that $x_k \in S(\bar{A}, b_k, \bar{\mu})$ is the (by assumption) unique solution of the problem $P(\bar{A}, b_k, \bar{\mu})$ for all $k \in \mathbb{N}$ sufficiently large. By [13, Theorem 3.1], which provides a full characterization of solution uniqueness for convex optimization problems, we obtain that

$$\ker \bar{A} \cap \mathrm{cone}(\partial g^*(\bar{z}) - x_k) = \{0\} \tag{53}$$

for all $k \in \mathbb{N}$ sufficiently large. As $x_k \in \mathrm{ri}\, \partial g^*(\bar{z})$, we have $\mathrm{cone}\,(\partial g^*(\bar{z}) - x_k) = \mathrm{par}\, \partial g^*(\bar{z})$ (cf. (9)). This together with (53) verifies the condition (45). $\square$

We are now in a position to establish the the main result of this paper, which shows that condition (45) is not only sufficient for Lipschitz stability of the solution mapping (30) around $(\bar{A}, \bar{b}, \bar{\mu})$, but also necessary provided that the function g^* is $\mathcal{C}^2$-cone reducible. Moreover, we show that the solution mapping S is single-valued and Lipschitz continuous around the point in question if and only if it is single-valued, i.e., the Lipschitz stability is automatic whenever the mapping is single-valued. This equivalence is interesting, as the Lipschitz stability is usually not a free property for single-valued mappings in general.

Theorem 3.8 (Full characterization of Lipschitz stability of solution mapping S). *Suppose that $\bar{x}$ is an optimal solution of problem $P(\bar{A}, \bar{b}, \bar{\mu})$ and that the function g^* is $\mathcal{C}^2$-cone reducible at $\bar{z} = \frac{1}{\bar{\mu}}\bar{A}^*(\bar{b} - \bar{A}\bar{x})$. Then the following are equivalent:*

  (i) *Condition (45) is satisfied.*
  (ii) *The solution mapping $S(A, b, \mu)$ in (30) of problem (29) is single-valued and Lipschitz continuous around $(\bar{A}, \bar{b}, \bar{\mu})$ with $\bar{x} = S(\bar{A}, \bar{b}, \bar{\mu})$.*
  (iii) *The solution mapping $S(\bar{A}, \cdot, \cdot)$ is single-valued and Lipschitz continuous around $(\bar{b}, \bar{\mu})$ with $\bar{x} = S(\bar{A}, \bar{b}, \bar{\mu})$.*
  (iv) *The solution mapping $S(\bar{A}, \cdot, \bar{\mu})$ is single-valued and Lipschitz continuous around $\bar{b}$ with $\bar{x} = S(\bar{A}, \bar{b}, \bar{\mu})$.*
  (v) *The solution mapping S in (30) is single-valued around $(\bar{A}, \bar{b}, \bar{\mu})$ with $\bar{x} = S(\bar{A}, \bar{b}, \bar{\mu})$.*
  (vi) *The solution mapping $S(\bar{A}, \cdot, \cdot)$ is single-valued around $(\bar{b}, \bar{\mu})$ with $\bar{x} = S(\bar{A}, \bar{b}, \bar{\mu})$.*
  (vii) *The solution mapping $S(\bar{A}, \cdot, \bar{\mu})$ is single-valued around $\bar{b}$ with $\bar{x} = S(\bar{A}, \bar{b}, \bar{\mu})$.*

Proof. The implication [**(i)**$\Rightarrow$**(ii)**] follows from Theorem 3.5. Implications [**(ii)**$\Rightarrow$**(iii)**$\Rightarrow$**(iv)**$\Rightarrow$**(vii)**] are trivial. We also have [**(ii)**$\Rightarrow$**(v)**$\Rightarrow$**(vi)**$\Rightarrow$**(vii)**]. Moreover, [**(vii)**$\Rightarrow$**(i)**] is obtained from Theorem 3.7. $\square$

Theorem 3.8 generalizes the recent results in [24, Theorem 3.12 and Theorem 4.7], which establish the equivalence between **(iii)** and **(vi)** when regularizers g include the ℓ_1/ℓ_2 norm and the nuclear norm via a different approach. Note from our Remark 3.6, conjugates of the latter two regularizers are $\mathcal{C}^2$-cone reducible. Moreover, our condition (45) recovers the characterizations of Lipschitz stability for the solution mapping in (30) used in [24] for these two cases and also the one in [3, Assumption 4.3] when g is the ℓ_1-norm, which appeared in [23, 34] as a sufficient condition for the Lasso problem. The implication [**(vii)**$\Longrightarrow$**(iv)**] in this theorem is also stronger than [11, Theorem 3.3], which shows that if $S(\bar{A}, \cdot, \bar{\mu})$ is single-valued around $\bar{b}$, then it is continuous when g is a convex piecewise linear function that is $\mathcal{C}^2$-cone reducible; see, e.g., Remark 3.6. Stability of the solution mapping S with respect to variable b seems to be important in the proof of our Theorem 3.7 and Theorem 3.8. By fixing the parameter $b = \bar{b}$, our condition (45) is also sufficient for the Lipschitz stability of solution mappings $S(\cdot, \bar{b}, \bar{\mu})$, $S(\cdot, \bar{b}, \cdot)$, and $S(\bar{A}, \bar{b}, \cdot)$ around the corresponding points due to the implication [**(i)**$\Rightarrow$**(ii)**] in Theorem 3.8. Whether it is necessary is an open question.

We close out this section by showing that the Lipschitz stability of the solution mapping of problem (29) is automatic when $\bar{x}$ is the unique solution of problem $P(\bar{A}, \bar{b}, \bar{\mu})$ and the so-called *Dual Strict Complementarity Condition* is satisfied and granted that g^* is $\mathcal{C}^2$-cone reducible.

Theorem 3.9. *Suppose that $\bar{x}$ is an optimal solution of problem $P(\bar{A}, \bar{b}, \bar{\mu})$ and that the function g^* is $\mathcal{C}^2$-cone reducible at $\bar{z} = \frac{1}{\bar{\mu}}\bar{A}^*(\bar{b} - \bar{A}\bar{x})$. Moreover, suppose that the following Dual Strict Complementarity Condition holds*

$$\bar{x} \in \operatorname{ri} \partial g^*(\bar{z}). \tag{54}$$

Then the solution mapping S in (30) is single-valued and Lipschitz continuous around $(\bar{A}, \bar{b}, \bar{\mu})$ with $S(\bar{A}, \bar{b}, \bar{\mu}) = \bar{x}$ if and only if $\bar{x}$ is the unique solution of problem $P(\bar{A}, \bar{b}, \bar{\mu})$.

Proof. The direction "$\Rightarrow$" is trivial. To prove the opposite implication, we suppose that $\bar{x}$ is the unique solution of problem $P(\bar{A}, \bar{b}, \bar{\mu})$ and that the condition (54) is satisfied. By [13, Theorem 3.1],

we have

$$\ker \bar{A} \cap \operatorname{cone}(\partial g^*(\bar{z}) - \bar{x}) = \{0\} \quad \text{with} \quad \bar{z} = -\frac{1}{\bar{\mu}} \bar{A}^*(\bar{A}\bar{x} - \bar{b}). \tag{55}$$

As $\bar{x} \in \operatorname{ri} \partial g^*(\bar{z})$ by (54), we obtain from (9) that

$$\operatorname{cone}(\partial g^*(\bar{z}) - \bar{x}) = \operatorname{par} \partial g^*(\bar{z}).$$

This together (55) tells us that condition (45) is satisfied. By Theorem 3.8, the solution mapping S is single-valued and Lipschitz continuous around $(\bar{A}, \bar{b}, \bar{\mu})$ with $S(\bar{A}, \bar{b}, \bar{\mu}) = \bar{x}$. $\qquad\square$

The above result resembles [24, Corollary 4.8] when g is the nuclear norm and the condition (54) is replaced by the *Strict Complementarity Condition*:

$$\bar{v} \in \operatorname{ri} \partial g(\bar{x}). \tag{56}$$

For the case of nuclear norm, it is shown in [13, Remark 3.7] that (54) and (56) are equivalent. But at this stage, we do not know their correlation for more general $\mathcal{C}^2$-cone reducible functions g^*.

4. FULL STABILITY OF CONVEX ADDITIVE-COMPOSITE PROBLEMS

In this section, we study the following optimization problem

$$\tilde{P}(\bar{p}) \qquad \min_{x \in \mathbb{X}} \quad f(x, \bar{p}) + g(x), \tag{57}$$

where $\mathbb{X}$ and $\mathbb{P}$ are two Euclidean spaces, $f : \mathbb{X} \times \mathbb{P} \to \mathbb{R} \cup \{+\infty\}$, and $g : \mathbb{X} \to \mathbb{R} \cup \{+\infty\}$ is closed, proper, convex. Suppose that $\bar{x} \in \operatorname{dom} f(\cdot, \bar{p}) \cap \operatorname{dom} g$ is an optimal solution of problem (57). Throughout this section, we assume that

   (A) $f(\cdot, \bar{p})$ is a convex function;
   (B) $f(\cdot, \cdot)$ is twice continuously differentiable around $(\bar{x}, \bar{p})$.

Define the function $\varphi : \mathbb{X} \times \mathbb{P} \to \mathbb{R} \cup \{+\infty\}$ by

$$\varphi(x, p) := f(x, p) + g(x) \quad \text{for all} \quad (x, p) \in \mathbb{X} \times \mathbb{P}. \tag{58}$$

The two-parametric perturbation of problem (57) is constructed by

$$\min_{x \in \mathbb{X}} \quad \varphi(x, p) - \langle v, x \rangle, \tag{59}$$

with *basic* perturbation $p \in \mathbb{P}$ and tilt parameter $v \in \mathbb{X}$. This certainly covers the problem (31).

Let us recall the definition of *full stability* introduced by Levy, Poliquin, and Rockafellar [19], which is a far-reaching extension of the notion of tilt stability by Poliquin and Rockafellar [25].

Definition 4.1 (Full stability and tilt stability). *The point $\bar{x}$ is called a* fully stable *optimal solution of problem* (57) *if there exists $\gamma > 0$ such that the solution map*

$$M_\gamma(v, p) := \operatorname{argmin} \{\varphi(x, p) - \langle v, x \rangle \,|\, x \in \mathbb{B}_\gamma(\bar{x})\} \quad \text{for} \quad (v, p) \in \mathbb{X} \times \mathbb{P} \tag{60}$$

and the mapping

$$m_\gamma(v, p) := \inf \{\varphi(x, p) - \langle v, x \rangle \,|\, x \in \mathbb{B}_\gamma(\bar{x})\} \tag{61}$$

are single-valued and Lipschitz continuous on some neighborhood of $(0, \bar{p}) \in \mathbb{X} \times \mathbb{P}$ with $M_\gamma(0, \bar{p}) = \bar{x}$.

The point $\bar{x}$ is called a tilt stable *optimal solution of problem* (57) *if the mapping $M_\gamma(\cdot, \bar{p})$ is single-valued and Lipschitz continuous on some neighborhood of $0 \in \mathbb{X}$ with $M_\gamma(0, \bar{p}) = \bar{x}$.*

As the function f is twice continuously differentiable around $(\bar{x}, \bar{p})$ and g is closed, proper, convex, the function φ is *continuously prox-regular* in x at $\bar{x}$ for 0, with compatible parameterization by p at $\bar{p}$ in the sense of [19, Definition 2.1]. Moreover, we observe that the so-called *basic constraint qualification*

$$(0, q) \in \partial^\infty \varphi(\bar{x}, \bar{p}) \quad \Longrightarrow \quad q = 0 \tag{62}$$

is also satisfied, where $\partial^\infty \varphi(\bar{x}, \bar{p})$ is the singular subdifferential of the function φ at $(\bar{x}, \bar{p})$ defined in (18). This is due to the sum rule of the singular subdifferential [20, Proposition 1.107] (see also [27, Corollary 10.9]), which gives

$$\partial^\infty \varphi(\bar{x}, \bar{p}) = \partial^\infty g(\bar{x}) \times \{0\}$$

using assumption (B). According to [19, Proposition 3.5], when the above basic constraint qualification is satisfied, the condition in the definition of full stability that $m_\gamma(\cdot, \cdot)$ is Lipschitz continuous on some neighborhood of $(0, \bar{p}) \in \mathbb{X} \times \mathbb{P}$ is automatically met, provided that $M_\gamma(0, \bar{p}) = \{\bar{x}\}$. It is also worth mentioning that the latter condition $M_\gamma(0, \bar{p}) = \bar{x}$ means that $\bar{x}$ is the unique (global) optimal solution of problem (57), since both functions $f(\cdot, \bar{p})$ and $g(\cdot)$ are convex in our framework.

Let us recall here [19, Theorem 2.3], [21, Theorem 4.4], and [25, Theorem 1.3], which provides characterizations of full stability and tilt stability via second-order analysis.

Theorem 4.2 (Characterizations of full stability). *Suppose that $\bar{x}$ is an optimal solution of problem* (57). *Then $\bar{x}$ is a fully stable optimal solution of problem* (57) *if and only if the following two conditions hold:*

(a) $\inf \{\langle z, w \rangle \mid (z, q) \in D^*(\partial_x \varphi)(\bar{x}, \bar{p} \mid 0)(w)\} > 0$ *for all $w \in \mathbb{X} \backslash \{0\}$.*
(b) $(0, q) \in D^*(\partial_x \varphi)(\bar{x}, \bar{p} \mid 0)(0) \implies q = 0$.

Moreover, $\bar{x}$ is a tilt stable optimal solution of problem (57) *if and only if*

$$\inf \{\langle z, w \rangle \mid z \in D^* \partial \varphi_{\bar{p}}(\bar{x} \mid 0)(w)\} > 0 \quad \text{for all} \quad w \in \mathbb{X} \backslash \{0\}, \tag{63}$$

where $\varphi_{\bar{p}}(\cdot) := \varphi(\cdot, \bar{p})$.

It is worth mentioning that in the original result of Levy, Poliquin, and Rockafellar [19, Theorem 2.4], the condition in the above part (a) does not include the "infimum". This equivalent "infimum" form coming from [21, Theorem 4.4] is useful for us in this paper at some point.

Note also that $\partial_x \varphi(x, p) = \nabla_x f(x, p) + \partial g(x)$. It follows from the coderivative sum rule [20, Theorem 1.62] that

$$D^*(\partial_x \varphi)(\bar{x}, \bar{p} \mid 0)(w) = (\nabla^2_{xx} f(\bar{x}, \bar{p})^* w, \nabla_{xp} f(\bar{x}, \bar{p})^* w) + D^* \partial g(\bar{x} \mid -\nabla_x f(\bar{x}, \bar{p}))(w) \times \{0\}. \tag{64}$$

This tells us that condition (b) in Theorem 4.2 is always true in our setting. Moreover, condition (a) in Theorem 4.2 turns to

$$\langle \nabla^2_{xx} f(\bar{x}, \bar{p}) w, w \rangle + \inf\{\langle z, w \rangle \mid z \in D^* \partial g(\bar{x} \mid -\nabla_x f(\bar{x}, \bar{p}))(w)\} > 0 \quad \text{for all} \quad w \in \mathbb{X} \backslash \{0\}, \tag{65}$$

which is also condition (63) again due to the the sum rule of the coderivative [20, Theorem 1.62]:

$$D^* \partial_x \varphi_{\bar{p}}(\bar{x} \mid 0)(w) = \nabla^2_{xx} f(\bar{x}, \bar{p})^* w + D^* \partial g(\bar{x} \mid -\nabla_x f(\bar{x}, \bar{p}))(w).$$

Hence, both full stability and tilt stability are equivalent for our problem (57) at the optimal solution $\bar{x}$ and they are characterized by condition (65).

As $f(\cdot, \bar{p})$ is a convex function, $\nabla^2_{xx} f(\bar{x}, \bar{p})$ is a positive semidefinite operator. Its *square root operator* $\bar{A} := (\nabla^2_{xx} f(\bar{x}, \bar{p}))^{\frac{1}{2}} \in \mathcal{L}(\mathbb{X}, \mathbb{X})$ with $\bar{A}^* \bar{A} = \nabla^2_{xx} f(\bar{x}, \bar{p})$ is unique and well-defined. Condition (65) is equivalent to

$$\|\bar{A}w\|^2 + \inf\{\langle z, w \rangle \mid z \in D^* \partial \tilde{g}(\bar{x} \mid 0)(w)\} > 0 \quad \text{for all} \quad w \in \mathbb{X} \backslash \{0\} \tag{66}$$

with $\tilde{g}(x) := g(x) + \langle \nabla_x f(\bar{x}, \bar{p}), x - \bar{x} \rangle$ for any $x \in \mathbb{X}$. It follows from Theorem 4.2 that the above condition is satisfied if and only if $\bar{x}$ is a tilt stable optimal solution of the following problem

$$\min_{x \in \mathbb{X}} \quad \frac{1}{2} \|\bar{A}x - \bar{b}\|^2 + \tilde{g}(x) \quad \text{with} \quad \bar{b} := \bar{A}\bar{x}. \tag{67}$$

This optimization problem is in the format of (29). The above observation allows us to establish a simple necessary condition of the tilt stability as below.

Theorem 4.3 (Necessary condition for tilt stability)**.** *Suppose that $\bar{x}$ is a tilt stable optimal solution of problem* (57)*. Then we have*

$$\ker \nabla_{xx}^2 f(\bar{x}, \bar{p}) \cap \operatorname{par} \partial g^*(\bar{z}) = \{0\} \quad with \quad \bar{z} := -\nabla_x f(\bar{x}, \bar{p}). \tag{68}$$

Proof. Suppose that $\bar{x}$ is a tilt stable optimal solution of problem (57). As discussed above, it is also a tilt stable optimal solution of problem (67). In particular, by first-order optimality conditions, $\bar{z} \in \partial g(\bar{x})$, i.e., $\bar{x} \in \partial g^*(\bar{z})$, thus (68) is well-defined. Therefore, there exists a sequence $\{x_k \in \operatorname{ri} \partial g^*(\bar{z})\} \to \bar{x}$. Define

$$v_k := \nabla_{xx}^2 f(\bar{x}, \bar{p})(x_k - \bar{x}) \to 0$$

and consider the linearly perturbed problem

$$\min_{x \in \mathbb{X}} \quad \psi_k(x) := \frac{1}{2}\|\bar{A}x - \bar{b}\|^2 + \tilde{g}(x) - \langle v_k, \, x \rangle \tag{69}$$

As $\bar{z} \in \partial g(x_k)$ and $\bar{A}^* \bar{A} = \nabla_{xx}^2 f(\bar{x}, \bar{p})$, we have

$$\partial \psi_k(x_k) = \bar{A}^*(\bar{A}x_k - \bar{A}\bar{x}) + \partial g(x_k) - \bar{z} - v_k = \nabla_{xx}^2 f(\bar{x}, \bar{p})(x_k - \bar{x}) + \partial g(x_k) - \bar{z} - v_k \ni 0.$$

Hence, x_k is an optimal solution of problem (69). Since $v_k \to 0$ and $\bar{x}$ is a tilt stable optimal solution of problem (67), x_k is the unique solution of problem (69) (for all $k \in \mathbb{N}$ sufficiently large). Define $\hat{g}_k(x) := \tilde{g}(x) - \langle v_k, x \rangle$ and obtain from [13, Theorem 3.1] that

$$\ker \bar{A} \cap \operatorname{cone}(\partial \hat{g}_k^*(-v_k) - x_k) = \{0\}, \tag{70}$$

which is indeed the characterization for solution uniqueness of problem (69). Note that

$$\hat{g}_k(x) = g(x) - \langle \bar{z}, \, x - \bar{x} \rangle - \langle v_k, \, x \rangle.$$

We have $\partial \hat{g}_k(x) = \partial g(x) - \bar{z} - v_k$. It follows that $\partial \hat{g}_k^*(-v_k) = \partial g^*(\bar{z})$. Since $x_k \in \operatorname{ri} \partial g^*(\bar{z})$, we obtain from (9) that

$$\operatorname{cone}(\partial \hat{g}^*(-v_k) - x_k) = \operatorname{cone}(\partial g^*(\bar{z}) - x_k) = \operatorname{par} \partial g^*(\bar{z}).$$

This together with (70) and the fact that $\ker \bar{A} = \ker \nabla_{xx}^2 f(\bar{x}, \bar{p})$ verifies (68). $\qquad\square$

When the Fenchel conjugate g^* is $\mathcal{C}^2$-cone reducible, we show next that condition (68) is also sufficient for full stability and tilt stability of problem (57).

Theorem 4.4 (Characterization of full stability and sufficient condition for Lipschitz stability)**.** *Suppose that $\bar{x}$ is an optimal solution of problem* (57) *and that the function g^* is $\mathcal{C}^2$-cone reducible at $\bar{z} = -\nabla_x f(\bar{x}, \bar{p}) \in \partial g(\bar{x})$. Then the following are equivalent:*

  (i) *$\bar{x}$ is a fully stable optimal solution of problem* (57)*.*
  (ii) *$\bar{x}$ is a tilt stable optimal solution of problem* (57)*.*
  (iii) *Condition* (68) *is satisfied.*

Consequently, if additionally $f(\cdot, p)$ is convex for any $p \in \mathbb{P}$ around $\bar{p}$, then the solution mapping

$$S(p) := \operatorname{argmin}\{f(x,p) + g(x)|\, x \in \mathbb{X}\} \tag{71}$$

is single-valued and Lipschitz continuous around $\bar{p}$ with $S(\bar{p}) = \bar{x}$ provided that condition (68) *is satisfied.*

Proof. Suppose that $\bar{x}$ is an optimal solution of problem (57). Note that $\bar{x}$ is also an optimal solution of problem (67). The equivalence between **(i)** and **(ii)** follows the arguments discussed after Theorem 4.2. Moreover, the implication [**(ii)**⇒**(iii)**] is obtained by Theorem 4.3. It remains to verify [**(iii)**⇒**(ii)**]. Indeed, suppose that condition (68) is satisfied and consider the following linearly perturbed version of (67):

$$\min_{x \in \mathbb{X}} \quad \frac{1}{2}\|\bar{A}x - \bar{b}\|^2 + \tilde{g}(x) - \langle v, \, x \rangle \tag{72}$$

with tilt parameter $v \in \mathbb{X}$. As $\partial \tilde{g}^*(0) = \partial g^*(\bar{z})$, condition (68) is equivalent to

$$\ker \bar{A} \cap \operatorname{par} \partial \tilde{g}^*(0) = \{0\}.$$

Since g^* is $\mathcal{C}^2$-cone reducible at $\bar{z}$, we infer that $\tilde{g}^* = g^*((\cdot) - \bar{z}) - \langle \bar{z}, \bar{x} \rangle$ is $\mathcal{C}^2$-cone reducible at 0. Applying Theorem 3.5 to problem (72) tells us that its solution mapping with variable v is single-valued and Lipschitz continuous around 0, i.e., $\bar{x}$ is a tilt stable solution of problem (72). By Theorem 4.2, we have

$$\|\bar{A}w\|^2 + \inf\{\langle z, w \rangle \mid z \in D^* \partial \tilde{g}(\bar{x} \mid 0)(w)\} > 0 \quad \text{for all} \quad w \in \mathbb{X}\backslash\{0\}$$

as in (66). Note also that $\|\bar{A}w\|^2 = \langle \bar{A}^* \bar{A}w, w \rangle = \langle \nabla_{xx}^2 f(\bar{x}, \bar{p})w, w \rangle$. The above condition is equivalent to (65), which is exactly the condition (65) or (63). By Theorem 4.2, $\bar{x}$ is a tilt stable optimal solution of problem (57), which means **(ii)** is satisfied.

Finally, suppose that the function $f(\cdot, p)$ is convex for any $p \in \mathbb{P}$ around $\bar{p}$ and condition (68) is satisfied. Hence, $\bar{x}$ is a fully stable optimal solution. It follows that there exists some $\gamma > 0$ such that the mapping

$$M_\gamma(v, p) := \operatorname{argmin}\{f(x, p) + g(x) - \langle v, x \rangle \mid x \in \mathbb{B}_\gamma(\bar{x})\}$$

is single-valued and Lipschitz continuous on some neighborhood of $\bar{p}$ with $M_\gamma(0, \bar{p}) = \bar{x}$. This allows us to find $\varepsilon > 0$ sufficiently small such that $M_\gamma(\{0\} \times \mathbb{B}_\varepsilon(\bar{p})) \subset \operatorname{int} \mathbb{B}_\gamma(\bar{x})$. Hence, $M_\gamma(0, p)$ for $p \in \mathbb{B}_\varepsilon(\bar{p})$ is the unique *local* optimal solution of

$$\min_{x \in \mathbb{X}} \quad f(x, p) + g(x).$$

As the function $f(\cdot, p)$ is convex for $p \in \mathbb{B}_\varepsilon(\bar{p})$, $M_\gamma(0, p)$ is also the unique global optimal solution. As $S(p) \in M_\gamma(0, p)$ for $p \in \mathbb{B}_\varepsilon(\bar{p})$, we have $S(p) = M_\gamma(0, p)$. Hence, $S(p)$ is single-valued and Lipschitz continuous on $\mathbb{B}_\varepsilon(\bar{p})$. The proof is completed. $\square$

Let us conclude this section by providing a simple example showing that condition (68) is not necessary for the Lipschitz stability of the solution mapping $S(p)$ in (71). This essentially highlights the difference between Theorem 3.8 and Theorem 4.4, where the function f in Theorem 3.8 has a special structure as a quadratic function.

Example 4.5. Consider the following simple optimization problem

$$\min_{x \in \mathbb{R}} \quad x^4 p, \tag{73}$$

where $p \in \mathbb{R}$ is a parameter around $\bar{p} = 1$, $f(x, p) = x^4 p$, and $g(x) = 0$. The solution set $S(p) = \{0\}$ is single-valued and Lipschitz continuous around $\bar{p}$. At $\bar{p}$, the solution of problem (73) is $\bar{x} = 0$. Note that $\bar{z} = -\nabla_x f(\bar{x}, \bar{p}) = 0$, $g^*(v) = \delta_0(v)$ for $v \in \mathbb{R}$, and $\nabla_{xx}^2 f(\bar{x}, \bar{p}) = 0$. It follows that $\partial g^*(\bar{z}) = \mathbb{R}$ and $\ker \nabla_{xx}^2 f(\bar{x}, \bar{p}) = \mathbb{R}$, which implies

$$\ker \nabla_{xx}^2 f(\bar{x}, \bar{p}) \cap \operatorname{par} \partial g^*(\bar{z}) = \mathbb{R}.$$

Conditions (68) fails in this case.

5. CONCLUSION

In this paper, we provide necessary and sufficient conditions for the Lipschitz stability of the solution mapping (30) of regularized least-squares optimization problem in the format of (29). When the Fenchel conjugate of the regularizer g is $\mathcal{C}^2$-cone reducible, we show that the Lipschitz stability of the solution mapping (30) is equivalent to our condition (45). This result recovers the corresponding findings in [3, 24] when the regularizer g is the ℓ_1 norm, the ℓ_1/ℓ_2 norm, or the nuclear norm.

One of the open questions we plan to investigate in the future is the computation of the Lipschitz modulus of the solution mapping when Lipschitz stability occurs. In theory, this Lipschitz modulus

may be obtained by calculating the coderivative of the solution mapping (30); see, e.g., [20, Theorem 4.10] and [27, Theorem 9.39]. However, this approach requires some involved computations of second-order structures of the regularizer g, which can be particularly challenging when g does not have a polyhedral structure. Finding the Lipschitz modulus without relying on second-order information of g, as in our approach, remains an open area of research that we wish to continue exploring.

Another interesting topic that we aim to study is the sensitivity analysis of the solution mapping (30) under our condition (45). Some initial results in this direction have been established recently in [3, 4, 5, 6, 7, 32]. However, [24, Example 3.14] provides an example where the solution mapping (30) is single-valued and Lipschitz continuous but not differentiable when g is the ℓ_1/ℓ_2 norm. Identifying the conditions under which the solution mapping is differentiable is one of our ongoing projects.

References

[1] Bonnans JF, Ramírez H (2005) Perturbation analysis of second-order cone programming problems. Math. Program. 104(2):205–227.

[2] Bonnans JF, Shapiro A (2000) Perturbation Analysis of Optimization Problems. Springer Series in Operations Research. (Springer-Verlag, New York).

[3] Berk A, Brugiapaglia S, Hoheisel T (2023) Lasso reloaded: a variational analysis perspective with applications to compressed sensing. SIAM J. Math. Data Sci. 5(4):1102–1129.

[4] Berk A, Brugiapaglia S, Hoheisel T (2024) Square Root LASSO: well-posedness, Lipschitz stability and the tuning trade off. SIAM J. Optim., 34(3):2609–2637

[5] Bôlte J, Pauwels E (2021) Conservative set valued fields, automatic differentiation, stochastic gradient methods and deep learning. Math. Program. 188:19–51.

[6] Bôlte J, Le T, Pauwels E, Silveti-Falls A (2021) Nonsmooth implicit differentiation for machine learning and optimization. Advances in Neural Information Processing Systems (NeurIPS 2021). 34:13537–13549.

[7] Bôlte J, E Pauwel E, Silveti-Falls A (2024) Differentiating nonsmooth solutions to parametric monotone inclusion problems. SIAM J. Optim., 34(1):71–97.

[8] Dontchev AL, Rockafellar RT (1996) Characterizations of strong regularity for variational inequalities over polyhedral convex sets. SIAM J. Optim. 6(4):1087–1105.

[9] Dontchev AL, Rockafellar RT (2014) 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).

[10] Cui Y, Ding C, Zhao X (2017) Quadratic growth conditions for convex matrix optimization problems associated with spectral functions. SIAM J. Optim. 27(4):2332–2355.

[11] Everink JM, Dong Y, Andersen MS (2024) The geometry and well-posedness of sparse regularized linear regression. arXiv:2409.03461v1.

[12] Fadili J, Nghia TTA, Phan DN (2023) Geometric characterizations of strong minima with applications to nuclear norm minimization problems. arXiv:2308.09224.

[13] Fadili J, Nghia TTA, Phan DN (2024): Solution uniqueness of convex optimization problems via the radial cone. arXiv:2401.5356054.

[14] Gfrerer H, Outrata J (2022) On (local) analysis of multifunctions via subspaces contained in graphs of generalized derivatives. J. Math. Anal. Appl. 508(2):125895.

[15] Hu C, Yao W, Zhang J (2024) Decomposition method for Lipschitz stability of general LASSO-type problems. arXiv:2407.18884.

[16] Klatte D, Kummer B (2002) Nonsmooth Equations in Optimization: Regularity, Calculus, Methods and Applications. (Kluwer).

[17] Kojima M (1980) Strongly stable stationary solutions in nonlinear programs. Analysis and Computation of Fixed Points Proceedings of a Symposium Conducted by the Mathematics Research Center. (University of Wisconsin–Madison). 93–138.

[18] Kojima M, Hirabiyashi R (1984) Continuous deformations of nonlinear programs. Math. Program. Stud. 21:150–198.

[19] Levy AB, Poliquin RA, Rockafellar RT (2000) Stability of locally optimal solutions. SIAM J. Optim. 10(2):580–604.

[20] Mordukhovich BS (2006) Variational Analysis and Generalized Differentiation. (Springer).

[21] Mordukhovich BS, Nghia TTA, Rockafellar RT (2015) Full stability in finite-dimensional optimization. Math. Oper. Res. 40(1):226–252.
[22] Meng K, Wu P, Yang X (2024) Lipschitz continuity of solution multifunctions of extended ℓ_1 regularization problems. arXiv:2406.16053.
[23] Mairal J, Yu B (2012) Complexity analysis of the lasso regularization path. In Proceedings of the 29th International Conference on Machine Learning. 1835–1842.
[24] Nghia TTA (2024) Geometric characterizations of Lipschitz stability for convex optimization problems. arXiv:2402.05215.
[25] Poliquin RA and Rockafellar RT (1998) Tilt stability of a local minimum. SIAM J. Optim. 8(2):287–299.
[26] Rockafellar RT (1970) Convex Analysis. (Princeton University Press, Princeton, New Jersey).
[27] Rockafellar RT and Wets RJB (1998) Variational Analysis. Grundlehren der Mathematischen Wissenschaften. (Springer-Verlag, Berlin).
[28] Robinson SM (1980) Strongly regular generalized equations. Math. Oper. Res. 5:43–62.
[29] Robinson SM (1984) Local structure of feasible sets in nonlinear programming, part II: Nondegeneracy. Math. Programming Stud. 22:217–230.
[30] Robinson SM (1987) Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity. Math. Programming Stud., 30:45–66.
[31] Shapiro A (2003) Sensitivity analysis of generalized equations. J. Math. Sciences, 115(4):2554–2565, 2003.
[32] Vaiter S, Deledalle C, Fadili J, Peyré G, Dossal C (2017) The degrees of freedom of partly smooth regularizers. Ann. Inst. Statist. Math. 69:791–832.
[33] Sun D (2006) The strong second-order sufficient condition and constraint nondegeneracy in nonlinear semidefinite programming and their implications. Math. Oper. Res. 31(4):761–776.
[34] Tibshirani RJ (2013) The Lasso problem and uniqueness. Electron. J. Stat. 7:1456–1490.

DEPARTMENT OF INDUSTRIAL ENGINEERING AND OPERATIONS RESEARCH, UNIVERSITY OF CALIFORNIA, BERKELEY, BERKELEY, CA 94720, USA
  Email address: yingcui@berkeley.edu

DEPARTMENT OF MATHEMATICS AND STATISTICS, MCGILL UNIVERSITY, MONTRÉAL, QC H3A0B9, CANADA
  Email address: tim.hoheisel@mcgill.ca

DEPARTMENT OF MATHEMATICS AND STATISTICS, OAKLAND UNIVERSITY, ROCHESTER, MI 48309, USA
  Email address: nttran@oakland.edu

DEPARTMENT OF APPLIED MATHEMATICS, THE HONG KONG POLYTECHNIC UNIVERSITY, HUNG HOM, KOWLOON, HONG KONG
  Email address: defeng.sun@polyu.edu.hk