Noname manuscript No.
(will be inserted by the editor)

Kurdyka-Łojasiewicz exponent via inf-projection

Peiran Yu · Guoyin Li · Ting Kei Pong

Communicated by James Renegar

Abstract Kurdyka-Łojasiewicz (KL) exponent plays an important role in estimating the convergence rate of many contemporary first-order methods. In particular, a KL exponent of $\frac{1}{2}$ for a suitable potential function is related to local linear convergence. Nevertheless, KL exponent is in general extremely hard to estimate. In this paper, we show under mild assumptions that KL exponent is preserved via inf-projection. Inf-projection is a fundamental operation that is ubiquitous when reformulating optimization problems via the lift-and-project approach. By studying its operation on KL exponent, we show that the KL exponent is $\frac{1}{2}$ for several important convex optimization models, including some semidefinite-programming-representable functions and some functions that involve C^2-cone reducible structures, under conditions such as strict complementarity. Our results are applicable to concrete optimization models such as group fused Lasso and overlapping group Lasso. In addition, for nonconvex models, we show that the KL exponent of many difference-of-convex functions can be derived from that of their natural majorant functions, and the KL exponent of the Bregman envelope of a function is the same as that of the function itself. Finally, we estimate the KL exponent of the sum of the

Peiran Yu
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China.
E-mail: peiran.yu@connect.polyu.hk

Guoyin Li
Department of Applied Mathematics, University of New South Wales, Sydney, Australia.
This author is partially supported by a Future fellowship from Australian Research Council (FT130100038) and a discovery project from Australian Research Council (DP190100555).
E-mail: g.li@unsw.edu.au

Ting Kei Pong
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China.
This author was supported partly by Hong Kong Research Grants Council PolyU153005/17p.
E-mail: tk.pong@polyu.edu.hk.

least squares function and the indicator function of the set of matrices of rank at most k.

Keywords First-order methods · Convergence rate · Kurdyka-Łojasiewicz inequality · Kurdyka-Łojasiewicz exponent · Inf-projection

Mathematics Subject Classification (2020) 90C06 · 90C22 · 90C25 · 90C26

1 Introduction

Many problems in machine learning, signal processing and data analysis involve large-scale nonsmooth nonconvex optimization problems. These problems are typically solved using first-order methods, which are noted for their scalability and ease of implementation. Commonly used first-order methods include the proximal gradient method and its variants, and splitting methods such as Douglas-Rachford splitting method and its variants; see the recent expositions [17, 42] and references therein for more detail. In the general nonconvex nonsmooth setting, convergence properties of the sequences generated by these algorithms are typically analyzed by assuming a certain potential function to have the so-called Kurdyka-Łojasiewicz (KL) property.

The KL property originates from the seminal Łojasiewicz inequality that bounds the function value deviation of a real-analytic function in terms of its gradient; see [38]. This inequality was extended to the case of C^1 subanalytic functions by Kurdyka in [31] using the notion of desingularizing function. An important breakthrough was made in [12, 13], where the Łojasiewicz inequality was further generalized to nonsmooth cases by using tools of modern variational analysis and semialgebraic geometry. This generalization significantly broadened the applicability of the aforementioned KL inequality to nonconvex settings, and it allowed us to perform convergence rate analysis for various important algorithms in nonsmooth optimization and subgradient dynamical systems.

The KL property[1] is satisfied by a large class of functions such as proper closed semi-algebraic functions; see, for example, [5]. It has been the main workhorse for establishing convergence of sequences generated by various first-order methods, especially in nonconvex settings [4–6, 15]. Moreover, when it comes to estimating *local convergence rate*, the so-called KL exponent plays a key role; see, for example, [4, Theorem 2], [27, Theorem 3.4] and [33, Theorem 3]. Roughly speaking, an exponent of $\alpha \in (0, \frac{1}{2}]$ of a suitable potential function corresponds to a linear convergence rate, while an exponent of $\alpha \in (\frac{1}{2}, 1)$ corresponds to a sublinear convergence rate. However, as noted in [40, Page 63, Section 2.1], explicit estimation of KL exponent for a given function is difficult in general. Nevertheless, due to its significance in convergence rate analysis, KL exponent computation has become an important research topic in recent years and some positive results have been obtained. For instance, we now know the KL exponent of the maximum of finitely many polynomials [32, Theorem 3.3]

[1] See Definition 2.1 for the precise definition.

and the KL exponent of a class of quadratic optimization problems with matrix variables satisfying orthogonality constraints [37]. In addition, it has been shown that the KL exponent is closely related to several existing and widely-studied error bound concepts such as the Hölder growth condition and the first-order error bound mentioned in [14, 41, 53];[2] see for example, [14, Theorem 5], [22, Theorem 3.7], [22, Proposition 3.8], [23, Corollary 3.6] and [34, Theorem 4.1]. Taking advantage of these connections, we now also know that convex models that satisfy the second-order growth condition have KL exponent $\frac{1}{2}$, so do models that satisfy the first-order error bound condition together with a mild assumption on the separation of stationary values; see the recent work [18,34,60] for concrete examples. This sets the stage for developing calculus rules for KL exponent in [34] to deduce the KL exponent of a function from functions with known KL exponents. For example, it was shown in [34, Corollary 3.1] that under mild conditions, if f_i is a KL function with exponent $\alpha_i \in [0, 1)$, $1 \leq i \leq m$, then the KL exponent of $\min_{1\leq i\leq m} f_i$ is given by $\max_{1\leq i\leq m} \alpha_i$. This was then used in [34, Section 5.2] for showing that the least squares loss with smoothly clipped absolute deviation (SCAD) [25] or minimax concave penalty (MCP) regularization [59] has KL exponent $\frac{1}{2}$.

In this paper, we will further explore this line of research and study how KL exponent behaves under the *inf-projection* operation: this is a significant generalization of the operation of taking the minimum of finitely many functions. Precisely, let $\mathbb{X}$ and $\mathbb{Y}$ be two finite dimensional Hilbert spaces and let $F : \mathbb{X} \times \mathbb{Y} \to \mathbb{R} \cup \{\infty\}$ be a proper closed function,[3] we call the function $f(x) := \inf_{y\in\mathbb{Y}} F(x, y)$ for $x \in \mathbb{X}$ an inf-projection of F. The name comes from the fact that the strict epigraph of f, defined as $\{(x, r) \in \mathbb{X} \times \mathbb{R} : \ f(x) < r\}$, is equal to the projection of the strict epigraph of F onto $\mathbb{X} \times \mathbb{R}$. Functions represented in terms of inf-projections arise naturally in sensitivity analysis as *value functions*; see, for example, [16, Chapter 3.2]. Inf-projection also appears when representing functions as optimal values of linear programming problems, or more generally, semidefinite programming (SDP) problems; see [29] for semidefinite-programming-representable (SDP-representable) functions. It is known that inf-projection preserves nice properties of F such as convexity [46, Proposition 2.22(a)]. In this paper, we show that, under mild assumptions, the KL exponent is also preserved under inf-projection. Based on this result and the ubiquity of inf-projection, we are then able to study KL exponents of various important convex and nonconvex models that were out of reach in previous studies. These include convex models such as a large class of SDP-representable functions, and some functions with C^2-cone reducible structures, as well as nonconvex models such as difference-of-convex functions and Bregman envelopes. These models are discussed in details in Section 3.1 with the general strategy for deducing their KL exponents outlined.

---

[2] This type of first-order error bound is sometimes called the Luo-Tseng error bound; see [34, 58].

[3] We refer the readers to Section 2 for relevant definitions.

The rest of the paper is organized as follows. We present necessary notation and preliminary materials in Section 2. The KL exponent under inf-projection is studied in Section 3, and we outline how the results can be applied to deducing KL exponents of some optimization models in Section 3.1. Section 4 is devoted to deriving KL exponents for various structured convex models, and in Section 5, we study KL exponents for several nonconvex models. Finally, some concluding remarks are given in Section 6.

2 Notation and preliminaries

In this paper, we use $\mathbb{X}$ and $\mathbb{Y}$ to denote two finite dimensional Hilbert spaces. We use $\langle \cdot, \cdot \rangle$ to denote the inner product of the underlying Hilbert space and use $\|\cdot\|$ to denote the associated norm. Moreover, for a linear map $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$, we use $\mathcal{A}^*$ to denote its adjoint. Next, we let $\mathbb{R}$ denote the set of real numbers and let $\mathbb{R}^n$ denote the set of n-tuples of real numbers. We also let $\mathbb{R}^{m\times n}$ denote the set of all $m \times n$ matrices. The (trace) inner product of two matrices A and $B \in \mathbb{R}^{m\times n}$ is defined as $\langle A, B\rangle := \mathrm{tr}(A^T B)$, where tr denotes the trace of a square matrix. The Fröbenius norm of a matrix $A \in \mathbb{R}^{m\times n}$ is denoted by $\|A\|_F$, which is defined as $\|A\|_F := \sqrt{\mathrm{tr}(A^T A)}$. Finally, the space of $n \times n$ symmetric matrices is denoted by $\mathcal{S}^n$, the cone of $n \times n$ positive semidefinite matrices is denoted by $\mathcal{S}^n_+$, and we write $X \succeq 0$ (resp., $X \succ 0$) to mean $X \in \mathcal{S}^n_+$ (resp., $X \in \mathrm{int}\,\mathcal{S}^n_+$, where $\mathrm{int}\,\mathcal{S}^n_+$ is the interior of $\mathcal{S}^n_+$).

For a set $\mathfrak{D} \subseteq \mathbb{X}$, we denote the distance from an $x \in \mathbb{X}$ to $\mathfrak{D}$ as $\mathrm{dist}(x, \mathfrak{D}) := \inf_{y\in\mathfrak{D}} \|x - y\|$. The closure (resp., interior) of $\mathfrak{D}$ is denoted by $\mathrm{cl}\,\mathfrak{D}$ (resp., $\mathrm{int}\,\mathfrak{D}$), and we use $B(x, r)$ to denote the closed ball centered at $x \in \mathbb{X}$ with radius $r > 0$, i.e., $B(x, r) := \{u \in \mathbb{X} :\ \|u - x\| \le r\}$. For a convex set $\mathfrak{C} \subseteq \mathbb{X}$, we denote its relative interior by $\mathrm{ri}\,\mathfrak{C}$, and use $\mathfrak{C}^\circ$ to denote its polar, which is defined as

$$\mathfrak{C}^\circ := \{z \in \mathbb{X} :\ \langle x, z\rangle \le 1 \text{ for all } x \in \mathfrak{C}\}.$$

Finally, the indicator function of a nonempty set $\mathfrak{D} \subseteq \mathbb{X}$ is denoted by $\delta_{\mathfrak{D}}$, which equals zero in $\mathfrak{D}$ and is infinity otherwise. We use $\sigma_{\mathfrak{D}}$ to denote its support function, which is defined as $\sigma_{\mathfrak{D}}(x) := \sup_{z\in\mathfrak{D}} \langle x, z\rangle$ for $x \in \mathbb{X}$.

For a mapping $\Theta : \mathbb{X} \to \mathbb{Y}$ that is continuously differentiable on $\mathbb{X}$, we use $D\Theta(x)$ to denote the derivative mapping of Θ at $x \in \mathbb{X}$: this is the linear map defined by

$$[D\Theta(x)]h := \lim_{t\to 0} \frac{\Theta(x + th) - \Theta(x)}{t} \quad \text{for all } h \in \mathbb{X}.$$

We denote the adjoint of the derivative mapping by $\nabla\Theta(x)$. This latter mapping is referred to as the gradient mapping of Θ at x. Then, following [47, Definition 3.1], we say that a closed set $\mathfrak{D} \subseteq \mathbb{X}$ is C^2-cone reducible at $\bar{w} \in \mathfrak{D}$ if there exist a closed convex pointed cone $K \subseteq \mathbb{Y}$, $\rho > 0$ and a mapping $\Theta : \mathbb{X} \to \mathbb{Y}$ that maps $\bar{w}$ to 0 and is twice continuously differentiable in $B(\bar{w}, \rho)$ with $D\Theta(\bar{w})$ being onto, such that

$$\mathfrak{D} \cap B(\bar{w}, \rho) = \{w : \Theta(w) \in K\} \cap B(\bar{w}, \rho).$$

We say that the set $\mathfrak{D}$ is C^2-cone reducible if, for all $\bar{w} \in \mathfrak{D}$, $\mathfrak{D}$ is C^2-cone reducible at $\bar{w}$. It is known that convex polyhedral sets, the positive semidefinite cone and the second-order cone are all C^2-cone reducible; see, for example, the discussion following [47, Definition 3.1]. Finally, following the discussion right after [18, Definition 6], we say that an extended-real-valued function is C^2-cone reducible if its epigraph is a C^2-cone reducible set, where the epigraph of an extended-real-valued function $f : \mathbb{X} \to [-\infty, \infty]$ is defined as $\operatorname{epi} f := \{(x,t) \in \mathbb{X} \times \mathbb{R} :\ f(x) \le t\}$.

An extended-real-valued function $f : \mathbb{X} \to [-\infty, \infty]$ is said to be proper if its domain $\operatorname{dom} f := \{x \in \mathbb{X} :\ f(x) < \infty\} \neq \emptyset$ and it is never $-\infty$. A proper function is closed if it is lower semicontinuous. For a proper function f, its regular subdifferential at $x \in \operatorname{dom} f$ is defined in [46, Definition 8.3] by

$$\hat{\partial} f(x) := \left\{ \zeta \in \mathbb{X} :\ \liminf_{z \to x, z \neq x} \frac{f(z) - f(x) - \langle \zeta, z - x \rangle}{\|z - x\|} \ge 0 \right\}.$$

The subdifferential of f at $x \in \operatorname{dom} f$ (which is also called the limiting subdifferential) is defined in [46, Definition 8.3] by

$$\partial f(x) := \left\{ \zeta \in \mathbb{X} :\ \exists x^k \xrightarrow{f} x,\ \zeta^k \to \zeta \text{ with } \zeta^k \in \hat{\partial} f(x^k) \text{ for each } k \right\};$$

here, $x^k \xrightarrow{f} x$ means both $x^k \to x$ and $f(x^k) \to f(x)$. Moreover, we set $\partial f(x) = \hat{\partial} f(x) = \emptyset$ for $x \notin \operatorname{dom} f$ by convention, and write $\operatorname{dom} \partial f := \{x \in \mathbb{X} :\ \partial f(x) \neq \emptyset\}$. It is known in [46, Exercise 8.8] that $\partial f(x) = \{\nabla f(x)\}$ if f is continuously differentiable at x. Moreover, when f is proper convex, the limiting subdifferential reduces to the classical subdifferential in convex analysis; see [46, Proposition 8.12]. Finally, for a nonempty closed set $\mathfrak{D}$, we define its normal cone at an $x \in \mathfrak{D}$ by $N_{\mathfrak{D}}(x) := \partial \delta_{\mathfrak{D}}(x)$. If $\mathfrak{D}$ is in addition convex, we define its tangent cone at $x \in \mathfrak{D}$ by $T_{\mathfrak{D}}(x) := [N_{\mathfrak{D}}(x)]^\circ$.

For a proper convex function f, its Fenchel conjugate is

$$f^*(u) := \sup_x \{\langle u, x \rangle - f(x)\};$$

moreover, it is known that the following equivalence holds (see [45, Theorem 23.5]):

$$u \in \partial f(x) \iff f(x) + f^*(u) = \langle x, u \rangle \iff f(x) + f^*(u) \le \langle x, u \rangle. \quad (2.1)$$

For a proper closed convex function f, its asymptotic (or recession) function f^∞ is defined by $f^\infty(d) := \liminf_{t \to \infty, d' \to d} \frac{f(td')}{t}$; see [7, Theorem 2.5.1]. Finally, for a proper function f, we say that it is level-bounded if, for each $\alpha \in \mathbb{R}$, the set $\{x :\ f(x) \le \alpha\}$ is bounded.

For a proper function $F : \mathbb{X} \times \mathbb{Y} \to \mathbb{R} \cup \{\infty\}$, following [46, definition 1.16], we say that F is level-bounded in y locally uniformly in x if for each $\bar{x} \in \mathbb{X}$ and $\alpha \in \mathbb{R}$ there is a neighborhood V of $\bar{x}$ such that the set $\{(x,y) \in \mathbb{X} \times \mathbb{Y} :\ x \in V \text{ and } F(x,y) \le \alpha\}$ is bounded. When a function F is level-bounded in y locally uniformly in x, its inf-projection $f(x) := \inf_y F(x,y)$ has the following properties, which can be found in [46]. We include the proof for the convenience of the readers.

Lemma 2.1 *Let $F : \mathbb{X} \times \mathbb{Y} \to \mathbb{R} \cup \{\infty\}$ be a proper closed function and define $f(x) := \inf_{y\in\mathbb{Y}} F(x,y)$ and $Y(x) := \operatorname{Arg\,min}_{y\in\mathbb{Y}} F(x,y)$ for $x \in \mathbb{X}$. Suppose F is level-bounded in y locally uniformly in x. Then the following statements hold:*

(i) *The function f is proper and closed, and the set $Y(x)$ is nonempty and compact for any $x \in \operatorname{dom}\partial f$.*

(ii) *For any $x \in \operatorname{dom}\partial f$, it holds that*

$$\partial f(x) \subseteq \bigcup_{y\in Y(x)} \{\xi \in \mathbb{X} \ : \ (\xi, 0) \in \partial F(x,y)\}. \tag{2.2}$$

(iii) *For any $\bar{x} \in \operatorname{dom}\partial f$, it holds that*

$$\limsup_{\operatorname{dom}\partial f \ni x \xrightarrow{f} \bar{x}} Y(x) \subseteq Y(\bar{x}); \tag{2.3}$$

(iv) *For any $\bar{x} \in \operatorname{dom}\partial f$ and any $\nu > 0$, there exists $\epsilon > 0$ such that*

$$\operatorname{dist}(y, Y(\bar{x})) \le \frac{\nu}{2}$$

whenever $y \in Y(x)$ with $x \in B(\bar{x},\epsilon) \cap \operatorname{dom}\partial f$ and $|f(x) - f(\bar{x})| < \epsilon$.

Proof Since F is proper, closed and level-bounded in y locally uniformly in x, we have from [46, Theorem 1.17] that f is proper and closed, and $Y(x)$ is a nonempty compact set whenever $x \in \operatorname{dom}\partial f$. Applying [46, Theorem 10.13], we conclude that (2.2) holds for any $x \in \operatorname{dom}\partial f$.

We now prove (iii) and (iv) respectively. For (iii), fix any $\bar{x} \in \operatorname{dom}\partial f$ and any y^* satisfying $y^* \in \limsup_{\operatorname{dom}\partial f \ni x \xrightarrow{f} \bar{x}} Y(x)$ and recall from [46, Section 5B] that $\limsup_{\operatorname{dom}\partial f \ni x \xrightarrow{f} \bar{x}} Y(x)$ is defined as

$$\left\{ y : \ \exists x^k \xrightarrow{f} \bar{x}, \ y^k \to y \text{ with } y^k \in Y(x^k) \text{ and } x^k \in \operatorname{dom}\partial f \text{ for each } k \right\}.$$

So, there exist $x^k \xrightarrow{f} \bar{x}$ with $x^k \in \operatorname{dom}\partial f$ and $y^k \to y^*$ such that $y^k \in Y(x^k)$ for all k. Then we have

$$F(\bar{x}, y^*) \overset{\text{(a)}}{\le} \liminf_k F(x^k, y^k) \overset{\text{(b)}}{=} \liminf_k f(x^k) \overset{\text{(c)}}{=} f(\bar{x}),$$

where (a) is due to the closedness of F, (b) holds because $y^k \in Y(x^k)$, and (c) holds because $x^k \xrightarrow{f} \bar{x}$. The above relation implies that $y^* \in Y(\bar{x})$. This proves (2.3).

Finally, for (iv), fix any $\bar{x} \in \operatorname{dom}\partial f$ and any $\nu > 0$. Since F is level-bounded in y locally uniformly in x, there exist $\tilde{\epsilon} > 0$ and a bounded set D so that whenever $x \in B(\bar{x}, \tilde{\epsilon}) \cap \operatorname{dom} \partial f$, we have $\{y : F(x, y) \le f(\bar{x}) + 1\} \subseteq D$. Thus, for any x satisfying $x \in B(\bar{x}, \tilde{\epsilon}) \cap \operatorname{dom} \partial f$ and $f(x) < f(\bar{x}) + 1$, we obtain

$$Y(x) = \{y : F(x, y) \le f(x)\} \subseteq \{y : F(x, y) \le f(\bar{x}) + 1\} \subseteq D. \tag{2.4}$$

Since (2.3) holds, by picking $\eta > 0$ so that $D \subseteq B(0, \eta)$ and following the proof of [46, Proposition 5.12(a)], we see that for this η, there exists $\epsilon \in (0, \min\{\tilde{\epsilon}, 1\})$ such that

$$Y(x) = Y(x) \cap D \subseteq Y(x) \cap B(0, \eta) \subseteq Y(\bar{x}) + B(0, \nu/2),$$

whenever $x \in B(\bar{x}, \epsilon) \cap \operatorname{dom} \partial f$ and $|f(x) - f(\bar{x})| < \epsilon$, where the first equality follows from (2.4) and the facts that $\epsilon < \tilde{\epsilon}$ and $\epsilon < 1$. This further implies that

$$\operatorname{dist}(y, Y(\bar{x})) \le \frac{\nu}{2}.$$

for any $y \in Y(x)$ with $x \in B(\bar{x}, \epsilon) \cap \operatorname{dom} \partial f$ and $|f(x) - f(\bar{x})| < \epsilon$. □

We next recall the Kurdyka-Łojasiewicz (KL) property and the notion of KL exponent; see [4–6, 31, 34, 38]. This property has been used extensively in analyzing convergence of first-order methods; see, for example, [4–6, 15, 56].

Definition 2.1 (Kurdyka-Łojasiewicz property and exponent) We say that a proper closed function $h : \mathbb{X} \to \mathbb{R} \cup \{\infty\}$ satisfies the Kurdyka-Łojasiewicz (KL) property at $\hat{x} \in \operatorname{dom} \partial h$ if there are $a \in (0, \infty]$, a neighborhood V of $\hat{x}$ and a continuous concave function $\varphi : [0, a) \to [0, \infty)$ with $\varphi(0) = 0$ such that

(i) φ is continuously differentiable on $(0, a)$ with $\varphi' > 0$ on $(0, a)$;
(ii) For any $x \in V$ with $h(\hat{x}) < h(x) < h(\hat{x}) + a$, it holds that

$$\varphi'(h(x) - h(\hat{x}))\operatorname{dist}(0, \partial h(x)) \ge 1. \tag{2.5}$$

If h satisfies the KL property at $\hat{x} \in \operatorname{dom} \partial h$ and the $\varphi(s)$ in (2.5) can be chosen as $\bar{c}\, s^{1-\alpha}$ for some $\bar{c} > 0$ and $\alpha \in [0, 1)$, then we say that h satisfies the KL property at $\hat{x}$ with exponent α.

A proper closed function h satisfying the KL property at every point in $\operatorname{dom} \partial h$ is said to be a KL function, and a proper closed function h satisfying the KL property with exponent $\alpha \in [0, 1)$ at every point in $\operatorname{dom} \partial h$ is said to be a KL function with exponent α.

KL functions is a broad class of functions which arise naturally in many applications. For instance, it is known that proper closed semi-algebraic functions are KL functions with exponent $\alpha \in [0, 1)$; see, for example, [5]. KL property is a key ingredient in many contemporary convergence analysis for first-order methods, and the KL exponent plays an important role in identifying *local convergence rate*; see, for example, [4, Theorem 2], [27, Theorem 3.4] and [33, Theorem 3]. In this paper, we will study how the KL exponent behaves under inf-projection, and use the rules developed to compute the KL exponents of various functions and to derive new calculus rules for KL exponent.

Before ending this section, we present two auxiliary lemmas. The first lemma concerns the uniformized KL property. It is a specialization of [15, Lemma 6] and explicitly involves the KL exponent.

Lemma 2.2 (Uniformized KL property with exponent) *Suppose that $h : \mathbb{X} \to \mathbb{R} \cup \{\infty\}$ is a proper closed function and let Ω be a nonempty compact set with $\Omega \subseteq \operatorname{dom} \partial h$. If h takes a constant value on Ω and satisfies the KL property at each point of Ω with exponent α, then there exist $\epsilon, a, c > 0$ such that*

$$\operatorname{dist}(0, \partial h(x)) \geq c\,(h(x) - h(\bar{x}))^{\alpha}$$

for any $\bar{x} \in \Omega$ and any x satisfying $h(\bar{x}) < h(x) < h(\bar{x}) + a$ and $\operatorname{dist}(x, \Omega) < \epsilon$.

Proof Replace the $\varphi_i(t)$ in the proof of [15, Lemma 6] by $c_i t^{1-\alpha}$ for some $c_i > 0$. The desired conclusion can then be proved analogously as in [15, Lemma 6]. □

The next lemma is a direct consequence of results in [50]; see [50, Theorem 3.3] and the discussion following [50, Eq. (1.4)] concerning the degree of singularity for semidefinite feasibility system.

Lemma 2.3 (Error bound for standard SDP problems under strict complementarity) *Let $C \in \mathcal{S}^d$, $\mathcal{A} : \mathcal{S}^d \to \mathbb{R}^m$ be a linear map, $b \in \operatorname{Range}(\mathcal{A})$ and define the function $G : \mathcal{S}^d \to \mathbb{R} \cup \{\infty\}$ by*

$$G(X) := \langle C, X \rangle + \delta_{\mathfrak{L}}(X),$$

where $\mathfrak{L} = \mathcal{A}^{-1}\{b\} \cap \mathcal{S}^d_+$. Suppose that $\mathcal{A}^{-1}\{b\} \cap \operatorname{int} \mathcal{S}^d_+ \neq \emptyset$ and there exists $\bar{X} \in \mathfrak{L}$ satisfying $0 \in \operatorname{ri} \partial G(\bar{X})$. Then for any bounded neighborhood $\mathfrak{U}$ of $\bar{X}$, there exists $c > 0$ such that for any $X \in \mathfrak{U} \cap \mathfrak{L}$,

$$\operatorname{dist}(X, \operatorname{Arg\,min} G) \leq c \left(G(X) - G(\bar{X})\right)^{\frac{1}{2}}.$$

Proof Observe that

$$\begin{aligned} 0 \in \operatorname{ri} \partial G(\bar{X}) &\overset{\text{(a)}}{=} C + \operatorname{ri} N_{\mathfrak{L}}(\bar{X}) \overset{\text{(b)}}{=} C + \operatorname{ri}\left(N_{\mathcal{A}^{-1}\{b\}}(\bar{X}) + N_{\mathcal{S}^d_+}(\bar{X})\right) \\ &\overset{\text{(c)}}{=} C + \operatorname{ri} N_{\mathcal{A}^{-1}\{b\}}(\bar{X}) + \operatorname{ri} N_{\mathcal{S}^d_+}(\bar{X}), \end{aligned} \tag{2.6}$$

where (a) follows from [46, Exercise 8.8], (b) follows from [45, Theorem 23.8] and the assumption $\mathcal{A}^{-1}\{b\} \cap \operatorname{int} \mathcal{S}^d_+ \neq \emptyset$, and (c) follows from [45, Corollary 6.6.2]. Since $N_{\mathcal{A}^{-1}\{b\}}(\bar{X}) = \operatorname{Range}(\mathcal{A}^*)$, we deduce further from (2.6) the existence of $\bar{y}$ satisfying

$$\mathcal{A}^*\bar{y} - C \in \operatorname{ri} N_{\mathcal{S}^d_+}(\bar{X}). \tag{2.7}$$

Next, since $0 \in \partial G(\bar{X})$, we have that $\bar{X} \in \operatorname{Arg\,min} G$ and thus

$$\operatorname{Arg\,min} G = \{W : \ \mathcal{A}W = b\} \cap \{W : \ \langle C, W \rangle = \inf G\} \cap \mathcal{S}^d_+ \neq \emptyset.$$

This together with (2.7) implies that the singularity degree of the semidefinite feasibility system $\left(\{W : \ \mathcal{A}W = b\} \cap \{W : \ \langle C, W \rangle = \inf G\}, \mathcal{S}^d_+\right)$ is one.

Combining this with [24, Theorem 2.3], we conclude that for any bounded neighborhood $\mathfrak{U}$ of $\bar{X}$, there exists $c_1 > 0$ such that for any $X \in \mathfrak{U} \cap \mathfrak{L}$,

$$\begin{aligned} \operatorname{dist}(X, \operatorname{Arg\,min} G) &\leq c_1 \sqrt{\operatorname{dist}\left(X, \{W:\ \mathcal{A}W = b\} \cap \{W : \langle C, W\rangle = \inf G\ \}\right)} \\ &\leq c\left(\langle C, X\rangle - \inf G\right)^{\frac{1}{2}} = c\left(G(X) - G(\bar{X})\right)^{\frac{1}{2}}, \end{aligned}$$

where the second inequality holds for some $c > 0$ thanks to the Hoffman error bound [26, Lemma 3.2.3]. This completes the proof. □

Remark 2.1 In the above lemma, the Slater's condition $\mathcal{A}^{-1}\{b\} \cap \operatorname{int} \mathcal{S}^d_+ \neq \emptyset$ together with the relative interior (ri) condition $0 \in \operatorname{ri} \partial G(\bar{X})$ implies that (2.7) holds. The condition (2.7) is widely used in the SDP literature and is often referred to as the strict complementarity condition; see [43, 48, 52] for detailed discussions. In particular, it is known that if strict complementarity condition (2.7) holds , then the singular degree of the associated semidefinite feasibility system is one (see [39, Proposition 7] or the discussion following [50, Eq. (1.4)]).

As we shall see in Section 4, this strict complementarity condition is crucial for deriving a KL exponent of $\frac{1}{2}$ for some SDP representable functions.

3 KL exponent via inf-projection

In this section, we study how the KL exponent behaves under inf-projection. Specifically, given a proper closed function $F : \mathbb{X} \times \mathbb{Y} \to \mathbb{R} \cup \{\infty\}$ with known KL exponent, we would like to deduce the KL exponent of $\inf_{y\in\mathbb{Y}} F(\cdot, y)$ under suitable assumptions.

Theorem 3.1 (KL exponent via inf-projection) *Let $F : \mathbb{X}\times\mathbb{Y} \to \mathbb{R}\cup\{\infty\}$ be a proper closed function and define $f(x) := \inf_{y\in\mathbb{Y}} F(x, y)$ and $Y(x) := \operatorname{Arg\,min}_{y\in\mathbb{Y}} F(x, y)$ for $x \in \mathbb{X}$. Suppose that the function F is level-bounded in y locally uniformly in x. Let $\alpha \in [0, 1)$ and $\bar{x} \in \operatorname{dom} \partial f$.[4] Suppose in addition the following conditions hold:*

- (i) *It holds that $\partial F(\bar{x}, \bar{y}) \neq \emptyset$ for all $\bar{y} \in Y(\bar{x})$.*
- (ii) *The function F satisfies the KL property with exponent α at every point in $\{\bar{x}\} \times Y(\bar{x})$.*

Then f satisfies the KL property at $\bar{x}$ with exponent α.

Proof Using the nonemptiness and compactness of $Y(\bar{x})$ given by Lemma 2.1(i), and the facts that $F(x, y) \equiv f(\bar{x})$ on $\Omega := \{\bar{x}\} \times Y(\bar{x}) \subseteq \operatorname{dom} \partial F$ and F satisfies the KL property with exponent α at every point in Ω, we deduce from Lemma 2.2 that there exist $\nu, a, c > 0$ such that

$$\operatorname{dist}\left(0, \partial F(x, y)\right) \geq c\left(F(x, y) - f(\bar{x})\right)^{\alpha} \tag{3.1}$$

[4] Here, f is a proper closed function, thanks to Lemma 2.1(i).

for any (x, y) satisfying

$$f(\bar{x}) < F(x,y) < f(\bar{x}) + a \quad \text{and} \quad \operatorname{dist}((x,y),\Omega) < \nu. \tag{3.2}$$

By decreasing a if necessary, without loss of generality, we may assume $a \in (0,1)$.

Next, using Lemma 2.1(iv), we see that there exists $\epsilon \in (0, \min\{\nu/2, a\})$ such that

$$\operatorname{dist}(y, Y(\bar{x})) \le \frac{\nu}{2}$$

whenever $y \in Y(x)$ with $x \in B(\bar{x},\epsilon) \cap \operatorname{dom}\partial f$ and $f(\bar{x}) < f(x) < f(\bar{x}) + \epsilon$. Hence, for any $x \in B(\bar{x},\epsilon) \cap \operatorname{dom}\partial f$ with $f(\bar{x}) < f(x) < f(\bar{x}) + \epsilon$ and any $y \in Y(x)$, we have

$$\operatorname{dist}((x,y),\Omega) \le \|x - \bar{x}\| + \operatorname{dist}(y, Y(\bar{x})) \le \epsilon + \frac{\nu}{2} < \nu,$$

where the last inequality follows from the choice of ϵ. The above relation together with the fact that $\epsilon < a$ shows that the relation (3.2) holds for any such x and any $y \in Y(x)$. Thus, using (3.1) we conclude that for any such x and any $y \in Y(x)$,

$$\begin{aligned}&\operatorname{dist}(0, \partial f(x)) = \operatorname{dist}\left(0, \begin{bmatrix}\partial f(x)\\ 0\end{bmatrix}\right) \ge \inf_{y \in Y(x)} \operatorname{dist}\left(0, \partial F(x,y)\right)\\ &\ge \inf_{y\in Y(x)} c\left(F(x,y) - f(\bar{x})\right)^{\alpha} = c\left(f(x) - f(\bar{x})\right)^{\alpha},\end{aligned}$$

where the first inequality follows from (2.2) and the last equality follows from the definition of $Y(x)$. This completes the proof. □

Theorem 3.1 can be viewed as a generalization of [34, Theorem 3.1], which studies the KL exponent of the minimum of finitely many proper closed functions with known KL exponents. Indeed, let f_i, $1 \le i \le m$, be proper closed functions. If we let $\mathbb{Y} = \mathbb{R}$ and define $F : \mathbb{X} \times \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ by

$$F(x,y) = \begin{cases} f_y(x) & \text{if } y = 1, 2, \ldots, m, \\ \infty & \text{otherwise,} \end{cases} \tag{3.3}$$

then it is not hard to see that this F is a proper closed function, and $\inf_{y\in\mathbb{R}} F(x,y) = \min_{1\le i\le m} f_i(x)$ for all $x \in \mathbb{X}$. Moreover, one can check directly from the definition that

$$\partial F(x,y) = \begin{cases} \partial f_y(x) \times \mathbb{R} & \text{if } y = 1, 2, \ldots, m, \\ \emptyset & \text{otherwise.} \end{cases} \tag{3.4}$$

Thus, we have the following immediate corollary of Theorem 3.1, which is a slight generalization of [34, Theorem 3.1] by dropping the continuity assumption on $\min_{1\le i\le m} f_i$.

Corollary 3.1 (KL exponent for minimum of finitely many functions) *Let f_i, $1 \le i \le m$, be proper closed functions, and define $f := \min_{1\le i\le m} f_i$. Let $\bar{x} \in \operatorname{dom}\partial f \cap \bigcap_{i\in I(\bar{x})} \operatorname{dom}\partial f_i$, where $I(\bar{x}) := \{i : f_i(\bar{x}) = f(\bar{x})\}$. Suppose that for each $i \in I(\bar{x})$, the function f_i satisfies the KL property at $\bar{x}$ with exponent $\alpha_i \in [0, 1)$. Then f satisfies the KL property at $\bar{x}$ with exponent $\alpha = \max\{\alpha_i : i \in I(\bar{x})\}$.*

Proof Define F as in (3.3). Then F is proper and closed, and $f(x) = \inf_{y\in\mathbb{R}} F(x, y)$. Moreover, $I(x) = Y(x) := \operatorname{Arg\,min}_{y\in\mathbb{R}} F(x, y)$. It is clear that this F is level-bounded in y locally uniformly in x. Moreover, in view of (3.4) and the assumption that $\bar{x} \in \bigcap_{i\in I(\bar{x})} \operatorname{dom}\partial f_i$, we see that $\partial F(\bar{x}, \bar{y}) \neq \emptyset$ whenever $\bar{y} \in Y(\bar{x})$. Finally, it is routine to show that F satisfies the KL property with exponent α_i at $(\bar{x}, i)$ for $i \in I(\bar{x})$. Thus, F satisfies the KL property with exponent $\alpha = \max\{\alpha_i : i \in I(\bar{x})\}$ on $\{\bar{x}\} \times I(\bar{x})$. The desired conclusion now follows from Theorem 3.1. $\square$

The next corollary can be proved similarly as [34, Corollary 3.1] by using Corollary 3.1 in place of [34, Theorem 3.1].

Corollary 3.2 *Let f_i, $1 \le i \le m$, be proper closed functions with $\operatorname{dom} f_i = \operatorname{dom}\partial f_i$ for all i, and define $f := \min_{1\le i\le m} f_i$. Suppose that for each i, the function f_i is a KL function with exponent $\alpha_i \in [0, 1)$. Then f is a KL function with exponent $\alpha = \max\{\alpha_i : 1 \le i \le m\}$.*

Finally, we show in the next corollary that one can relax some conditions of Theorem 3.1 when F is in addition convex.

Corollary 3.3 (KL exponent via inf-projections under convexity) *Let $F : \mathbb{X} \times \mathbb{Y} \to \mathbb{R} \cup \{\infty\}$ be a proper closed convex function and define $f(x) := \inf_{y\in\mathbb{Y}} F(x, y)$ and $Y(x) := \operatorname{Arg\,min}_{y\in\mathbb{Y}} F(x, y)$ for $x \in \mathbb{X}$. Suppose there exists $\bar{u}$ such that $f(\bar{u}) \in \mathbb{R}$ and $Y(\bar{u})$ is nonempty and compact. Then the following statements hold:*

(i) *The function f is proper and closed, and $Y(x)$ is nonempty and compact for any $x \in \operatorname{dom}\partial f$.*
(ii) *It holds that $\partial F(x, y) \neq \emptyset$ for all $x \in \operatorname{dom}\partial f$ and $y \in Y(x)$.*
(iii) *If $\bar{x} \in \operatorname{dom}\partial f$, $\alpha \in [0, 1)$ and the function F satisfies the KL property with exponent α at every point in $\{\bar{x}\} \times Y(\bar{x})$, then f satisfies the KL property at $\bar{x}$ with exponent α.*

Proof For (i), we first show that F is level-bounded in y locally uniformly in x. Suppose to the contrary that there exist $x_0 \in \mathbb{X}$ and $\beta \in \mathbb{R}$ so that $\mathfrak{C} := \{(x, y) : x \in B(x_0, 1) \text{ and } F(x, y) \le \beta\}$ is unbounded. Then there exists $\{(x^k, y^k)\} \subset \mathfrak{C}$ with $\|y^k\| \to \infty$. By passing to a subsequence if necessary, we may assume $\lim_{k\to\infty} \frac{y^k}{\|y^k\|} = d$ for some d with $\|d\| = 1$. Since $F(x^k, y^k) \le \beta$ and $\{x^k\} \subset B(x_0, 1)$ is bounded, we have

$$F^\infty(0, d) \le \liminf_{k\to\infty} \frac{F(x^k, y^k)}{\|(x^k, y^k)\|} \le \liminf_{k\to\infty} \frac{\beta}{\|(x^k, y^k)\|} = 0,$$

where F^∞ is the asymptotic function of F and the first inequality follows from [7, Theorem 2.5.1]. This together with the convexity of F and [7, Proposition 2.5.2] shows that

$$F(x, y + td) \leq F(x, y) \quad \text{for all } t > 0 \text{ and for all } (x, y) \in \operatorname{dom} F.$$

Since $Y(\bar{u}) \neq \emptyset$ and $f(\bar{u}) \in \mathbb{R}$, we have $\{\bar{u}\} \times Y(\bar{u}) \subseteq \operatorname{dom} F$. Hence, we can take $\bar{v} \in Y(\bar{u})$ and set $x = \bar{u}$ and $y = \bar{v}$ in the above display to conclude that $F(\bar{u}, \bar{v} + td) \leq F(\bar{u}, \bar{v})$ for all $t > 0$. This further implies that $\bar{v} + td \in Y(\bar{u})$ for all $t > 0$, which contradicts the compactness of $Y(\bar{u})$. Thus, for any $x_0 \in \mathbb{X}$ and $\beta \in \mathbb{R}$, the set $\{(x, y) : x \in B(x_0, 1) \text{ and } F(x, y) \leq \beta\}$ is bounded. Using Lemma 2.1(i), we see that (i) holds.

Next, we prove (ii). To this end, fix any $u \in \operatorname{dom} \partial f$ and $v \in Y(u)$. Note that the function f is convex as inf-projection of the convex function F; see [46, Proposition 2.22(a)]. Now, for the proper convex function f, we have from the definition that $f^*(w) = \sup_x\{\langle w, x\rangle - f(x)\} = \sup_{x,y}\{\langle w, x\rangle - F(x, y)\} = F^*(w, 0)$ for any $w \in \mathbb{X}$. Taking a $\bar{w} \in \partial f(u)$ and using (2.1), we see further that for any $v \in Y(u)$,

$$F(u, v) + F^*(\bar{w}, 0) = f(u) + f^*(\bar{w}) = \langle u, \bar{w}\rangle,$$

where the equality $F(u, v) = f(u)$ holds because $v \in Y(u)$. In view of (2.1), the above relation further implies that $(\bar{w}, 0) \in \partial F(u, v)$. This proves (ii).

Now, suppose in addition that $\bar{x} \in \operatorname{dom} \partial f$, $\alpha \in [0, 1)$ and the function F satisfies the KL property with exponent α at every point in $\{\bar{x}\} \times Y(\bar{x})$. Recall that we have shown that F is level-bounded in y locally uniformly in x in the proof of item (i) and we have $\{\bar{x}\} \times Y(\bar{x}) \subseteq \operatorname{dom} \partial F$ from item (ii). The conclusion (iii) now follows by applying Theorem 3.1. □

Remark 3.1 In addition to the inf-projection, another closely related operation, which appears frequently in optimization, would be taking the supremum over a family of functions. However, we would like to point out that, as opposed to the inf-projection, the supremum operation may not preserve KL exponents. For example, consider $F : \mathbb{R}^2 \to \mathbb{R}$ defined by $F = \max\{f_1, f_2\}$ with $f_1(x) = x_1^2$ and $f_2(x) = (x_1 + 1)^2 + x_2^2 - 1$. Clearly, f_1 and f_2 are both quadratic and are KL functions with exponent $\frac{1}{2}$. On the other hand, it was shown in [30, Page 1617] that F has an optimal solution at $(0, 0)$ and the KL exponent of F at $(0, 0)$ is $\frac{3}{4}$ and cannot be $\frac{1}{2}$. It would be of interest to see, under what additional conditions, the supremum operation can preserve the KL exponents. This could be one interesting future research direction.

3.1 Optimization models that can be written as inf-projections

Inf-projection is ubiquitous in optimization. In this section, we present some commonly encountered models that can be written as inf-projections. This includes a large class of semidefinite-programming-representable (SDP-representable) functions, rank constrained least squares problems, and Bregman envelopes. These are important convex and nonconvex models whose explicit KL exponents were out of reach in previous studies. In Sections 4 and 5, we will study their KL exponents based on their inf-projection representations, Theorem 3.1 and Corollary 3.3.

3.1.1 Convex models that can be written as inf-projections

(i) SDP-representable functions Following [29, Eq. (1.3)], we say that a function $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, is semidefinite-programming-representable (SDP-representable) if its epigraph can be expressed as the feasible region of some SDP problems, i.e., epi f equals

$$\left\{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \ \exists u \in \mathbb{R}^N \ \text{ s.t. } A_{00} + A_0 t + \sum_{i=1}^n A_i x_i + \sum_{j=1}^N B_j u_j \succeq 0\right\} \tag{3.5}$$

for some $\{A_{00}, A_0, A_1, \ldots, A_n, B_1, \ldots, B_N\} \subset \mathcal{S}^d$, $d \geq 1$ and $N \geq 1$. These functions arise in various applications and include important examples such as least squares loss functions, ℓ_1 norm, and nuclear norm, etc; see, for example, [11, Section 4.2] for more discussions. Using the symmetric matrices in (3.5), we define a linear map $\mathcal{A} : \mathcal{S}^d \to \mathbb{R}^{n+N+1}$ as

$$\mathcal{A}(W) := [\langle A_1, W\rangle \ \cdots \ \langle A_n, W\rangle \ \langle B_1, W\rangle \ \cdots \ \langle B_N, W\rangle \ \langle A_0, W\rangle]^T. \tag{3.6}$$

Then it is routine to show that $\mathcal{A}^* : \mathbb{R}^{n+N+1} \to \mathcal{S}^d$ is given by $\mathcal{A}^*(x,u,t) = A_0 t + \sum_{i=1}^n A_i x_i + \sum_{j=1}^N B_j u_j$ for $(x,u,t) \in \mathbb{R}^n \times \mathbb{R}^N \times \mathbb{R}$. Now, if we define

$$\begin{aligned} &F(x,u,t) := t + \delta_{\mathfrak{D}}(x,u,t) \\ &\text{with } \ \mathfrak{D} = \{(x,u,t) : \ A_{00} + \mathcal{A}^*(x,u,t) \succeq 0\}, \end{aligned} \tag{3.7}$$

then it holds that $f(x) = \inf_{u,t} F(x,u,t)$ for all $x \in \mathbb{R}^n$. We will show in Theorem 4.1 (using Corollary 3.3) that a proper closed SDP-representable function has KL property with exponent $\frac{1}{2}$ at points satisfying suitable assumptions on the SDP representation of F in (3.7).

(ii) Sum of LMI-representable functions We say that a function $h : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, is LMI-representable (see [29, Eq. (1.1)]) if there exist symmetric matrices A_{00}, A_j, $j = 0, \ldots, n$, such that

$$\operatorname{epi} h = \left\{(x,t) \in \mathbb{R}^n \times \mathbb{R} : \ A_{00} + \sum_{j=1}^n A_j x_j + A_0 t \succeq 0\right\}.$$

It is clear that LMI-representable functions form a special class of SDP-representable functions. Many commonly used functions are LMI-representable such as the least squares loss function, the ℓ_1, ℓ_2, ℓ_∞ norm functions, the indicator functions of their corresponding norm balls, and the indicator function of the matrix operator norm ball, etc.

Let $f = \sum_{i=1}^m f_i$ be the sum of m proper closed LMI-representable functions. In Theorem 4.2, we show that f has KL property with exponent $\frac{1}{2}$ at points under suitable assumptions. Different from Theorem 4.1, which imposes the "strict complementarity condition" on the corresponding F in (3.7), Theorem 4.2 *directly* imposes such kind of condition on the original function f. Explicit optimization models which can be written as sum of LMI-representable functions include (non-overlapping) group Lasso and group fused Lasso, and are discussed in Example 4.1.

(iii) Sum of LMI-representable functions and the nuclear norm The nuclear norm has been used for inducing low rank of solutions in various applications; see, for example, [44] for more discussions. Noticing that the nuclear norm is a special SDP-representable function, we further consider the sum of LMI-representable functions and the nuclear norm:

$$f(X) := \sum_{k=1}^{p} f_k(X) + \|X\|_*, \tag{3.8}$$

where $X \in \mathbb{R}^{m\times n}$, $\|X\|_*$ denotes the nuclear norm of X (the sum of all singular values of X) and each $f_k : \mathbb{R}^{m\times n} \to \mathbb{R} \cup \{\infty\}$ is a proper closed LMI-representable function. Define a function $F : \mathcal{S}^{n+m} \to \mathbb{R} \cup \{\infty\}$ by

$$F(Z) := \sum_{k=1}^{p} f_k(X) + \frac{1}{2}(\mathrm{tr}(U) + \mathrm{tr}(V)) + \delta_{\mathcal{S}_+^{m+n}}(Z); \tag{3.9}$$

here, we partition the matrix variable $Z \in \mathcal{S}^{n+m}$ as follows:

$$Z = \begin{bmatrix} U & X \\ X^T & V \end{bmatrix}, \tag{3.10}$$

where $U \in \mathcal{S}^m$, $V \in \mathcal{S}^n$ and $X \in \mathbb{R}^{m\times n}$. Then one can show that $f(X) = \inf_{U,V} F(Z)$; see (4.26) below. In Theorem 4.3, we will show that f in (3.8) satisfies KL property with exponent $\frac{1}{2}$ at points $\bar{X}$ such that $0 \in \mathrm{ri}\,\partial f(\bar{X})$, under mild conditions. Explicit optimization models of the form (3.8) are introduced in Remark 4.3.

(iv) Convex models with C^2-cone reducible structure SDP representable functions are all semi-algebraic. As an attempt to go beyond semi-algebraicity, we analyze functions involving C^2-cone reducible structure. Specifically, we consider the following function $f : \mathbb{X} \to \mathbb{R} \cup \{\infty\}$:

$$f(x) := \ell(\mathcal{A}x) + \langle v, x\rangle + \gamma(x), \tag{3.11}$$

where γ is a closed gauge[5] whose polar gauge[6] is C^2-cone reducible, the function $\ell : \mathbb{Y} \to \mathbb{R}$ is strongly convex on any compact convex set and has locally Lipschitz gradient, $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ is a linear map, and $v \in \mathbb{X}$.

[5] A gauge is a nonnegative positively homogeneous convex function that vanishes at the origin.

[6] See [28, Proposition 2.1(iii)].

Notice that $f(x) = \inf_t F(x,t)$, where

$$F(x,t) := \ell(\mathcal{A}x) + \langle v, x\rangle + t + \delta_{\mathfrak{D}}(x,t), \tag{3.12}$$

with $\mathfrak{D} = \{(x,t) \in \mathbb{X} \times \mathbb{R} : \gamma(x) \le t\}$. In Section 4.4, we will deduce that f in (3.11) has KL property with exponent $\frac{1}{2}$ at points satisfying assumptions involving relative interior of some subdifferential sets; see Corollary 4.1. Optimization models in the form of (3.11) are presented in Example 4.2.

3.1.2 Nonconvex optimization models that can be written as inf-projections

(i) Difference-of-convex functions We consider difference-of-convex (DC) functions of the following form:

$$f(x) = P_1(x) - P_2(\mathcal{A}x), \tag{3.13}$$

where $P_1 : \mathbb{X} \to \mathbb{R} \cup \{\infty\}$ is a proper closed convex function, $P_2 : \mathbb{Y} \to \mathbb{R}$ is a continuous convex function and $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ is a linear map. These functions arise in many contemporary applications including compressed sensing; see, for example, [1, 51, 56, 57] and references therein. In the literature, the following function is a typically used majorant for designing and analyzing algorithms for minimizing DC functions. It is obtained from (3.13) by majorizing the concave function $-P_2$ using the Fenchel conjugate P_2^* of P_2:

$$F(x,y) = P_1(x) - \langle \mathcal{A}x, y\rangle + P_2^*(y). \tag{3.14}$$

Note that $f(x) = \inf_y F(x,y)$ thanks to the definition of Fenchel conjugate and [45, Theorem 12.2]. In Theorem 5.1, we will deduce the KL exponent of f in (3.13) from that of F in (3.14).

(ii) Bregman envelope The Bregman envelope of a proper closed function $f : \mathbb{X} \to \mathbb{R} \cup \{\infty\}$, is defined in [10] as follows:

$$F_\phi(x) := \inf_y \{f(y) + \mathfrak{B}_\phi(y,x)\} \tag{3.15}$$

where $\phi : \mathbb{X} \to \mathbb{R}$ is a differentiable convex function and

$$\mathfrak{B}_\phi(y,x) = \phi(y) - \phi(x) - \langle \nabla\phi(x), y - x\rangle \tag{3.16}$$

is the Bregman distance. Note that F_ϕ is an inf-projection by definition. In Section 5.2, we will show that if ϕ satisfies Assumption 5.1 and f is a KL function with exponent $\alpha \in (0,1]$ and satisfies $\inf f > -\infty$, then F_ϕ in (3.15) is also a KL function with exponent $\alpha \in (0,1]$. As we shall see in Remark 5.1, the F_ϕ with ϕ satisfying Assumption 5.1 covers the widely studied Moreau envelope (see, for example, [46, Section 1G]) and the recently proposed forward-backward envelope [49].

(iii) Least squares loss function with rank constraint Consider the following least squares loss function with rank constraint:

$$f(X) := \frac{1}{2}\|\mathcal{A}X - b\|^2 + \delta_{\mathrm{rank}(\cdot)\le k}(X), \tag{3.17}$$

where $X \in \mathbb{R}^{m\times n}$, $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a linear map, $b \in \mathbb{R}^p$ and k is an integer between 1 and $\min\{m,n\} - 1$. The model above is considered in many applications such as principal components analysis (PCA); see [54] for more details. Notice that f in (3.17) is an inf-projection in the following form:

$$f(X) = \inf_U \left\{\frac{1}{2}\|\mathcal{A}X - b\|^2 + \delta_{\widehat{\mathfrak{D}}}(X,U)\right\}, \tag{3.18}$$

where

$$\widehat{\mathfrak{D}} := \{(X,U) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times(m-k)} :\ U^TX = 0 \text{ and } U^TU = I_{m-k}\},$$

and I_{m-k} is the identity matrix of size $m-k$. In Section 5.3, we first establish an auxiliary KL calculus rule concerning Lagrangian in Theorem 5.3. Then, using this result together with Theorem 3.1, we give an explicit KL exponent (dependent on n, m and k) of f in (3.17) in Theorem 5.4.

4 KL exponents for some convex models

4.1 Convex models with SDP-representable structure

In this section, we explore the KL exponent of SDP-representable functions introduced in Section 3.1.1(i). More specifically, we will deduce the KL exponent of a proper closed function f with its epigraph represented as in (3.5), under suitable conditions on F in (3.7). To this end, we collect the u components in $\mathfrak{D}$ in (3.7) for each fixed $x \in \mathrm{dom}\,\partial f$ and define the following set:

$$\mathfrak{D}_x = \left\{u \in \mathbb{R}^N :\ (x,u,f(x)) \in \mathfrak{D}\right\}. \tag{4.1}$$

Roughly speaking, these are extra variables that correspond to the "x-slice" in the "lifted" SDP representation. As we shall see in the proof of Theorem 4.2, when f is the sum of LMI-representable functions (which is SDP-representable), one can have $\mathfrak{D}_x = \{(f_1(x),\ldots,f_m(x))\}$.

We begin with three auxiliary lemmas. The first one relates the KL exponent of f, whose epigraph is represented as in (3.5), to that of F in (3.7).

Lemma 4.1 *Let $f : \mathbb{R}^n \to \mathbb{R}\cup\{\infty\}$ be a proper closed SDP-representable function with its epigraph represented as in* (3.5)*. Then the function F defined in* (3.7) *is proper, closed and convex.*

Next, suppose in addition that $\bar{x} \in \mathrm{dom}\,\partial f$, $\alpha \in [0,1)$, and that the following conditions hold:

(i) *The set $\mathfrak{D}_{\bar{x}}$ defined as in* (4.1) *is nonempty and compact.*

(ii) *The function F defined in* (3.7) *satisfies the KL property with exponent α at every point in $\{\bar x\}\times\mathfrak{D}_{\bar x}\times\{f(\bar x)\}$.*

Then f satisfies the KL property at $\bar x$ with exponent α.

Proof Observe from the definition that

$$f(x)=\inf_{u,t}F(x,u,t).$$

First, note that $\mathfrak{D}\neq\emptyset$ because f is proper. Since $\mathfrak{D}$ is clearly closed and convex, we conclude that F is proper, closed and convex. We will now check the conditions in Corollary 3.3 and apply the corollary to deduce the KL property of f from that of F.

To this end, by assumption, we see that F satisfies the KL property with exponent α on $\{\bar x\}\times\mathfrak{D}_{\bar x}\times\{f(\bar x)\}=\{\bar x\}\times\operatorname{Arg\,min}_{u,t}F(\bar x,u,t)$ and that $\mathfrak{D}_{\bar x}$ is nonempty and compact. The desired conclusion now follows from a direct application of Corollary 3.3. This completes the proof. □

The second lemma relates the KL exponent of F in (3.7) to that of another SDP-representable function with carefully constructed matrices involved in its representation.

Lemma 4.2 *Let f be a proper closed function and $\bar x\in\operatorname{dom}f$. Suppose that f is SDP-representable with its epigraph represented as in* (3.5)*, and that there exists (x^s,u^s,t^s) such that $A_{00}+\mathcal{A}^*(x^s,u^s,t^s)\succ 0$, where A_{00} and $\mathcal{A}$ are given in* (3.5) *and* (3.6) *respectively. Let F be defined as in* (3.7) *and $\mathfrak{D}_{\bar x}$ be defined as in* (4.1).[7] *Let $\bar u\in\mathfrak{D}_{\bar x}$ and suppose that $0\in\partial F(\bar x,\bar u,f(\bar x))$. Then the following statements hold:*

(i) *It holds that $A_0\neq 0$. Moreover, the set* $\operatorname{span}\{A_1,\dots,A_n,B_1,\dots,B_N,A_0\}$ *has an orthogonal basis $\{\hat A_0,\dots,\hat A_p\}$, where $p\ge 0$ and $\hat A_0\neq 0$, such that*

$$\begin{bmatrix}\mathbf{a}_1 & \dots & \mathbf{a}_n & \mathbf{b}_1 & \dots & \mathbf{b}_N & \mathbf{a}_0\end{bmatrix}=\begin{bmatrix}\hat{\mathbf{a}}_1\dots\hat{\mathbf{a}}_p & \hat{\mathbf{a}}_0\end{bmatrix}U$$

for some $U\in\mathbb{R}^{(p+1)\times(n+N+1)}$ having full row rank and the entries of the $(p+1)^{\text{th}}$ row of U are 0 except for $U_{p+1,n+N+1}=1$; here, $\mathbf{a}_i$, $\mathbf{b}_j$ and $\hat{\mathbf{a}}_k\in\mathbb{R}^{d^2}$ are the columnwise vectorization of the matrices A_i, B_j and $\hat A_k$, respectively.

(ii) *Define $F_1:\mathbb{R}^{p+1}\to\mathbb{R}\cup\{\infty\}$ by*

$$\begin{aligned}&F_1(z,t):=t+\delta_{\mathfrak{D}_1}(z,t)\\&\textit{with } \mathfrak{D}_1=\left\{(z,t):\ A_{00}+\hat A_0 t+\sum_{w=1}^{p}\hat A_w z_w\succeq 0\right\},\end{aligned}\tag{4.2}$$

where $p\ge 0$ and $\{\hat A_0,\dots,\hat A_p\}$ is the orthogonal basis constructed in (i).[8] *Suppose that $U(\bar x,\bar u,f(\bar x))\in\operatorname{dom}\partial F_1$ and F_1 satisfies the KL property*

[7] Notice that F is proper and closed thanks to the existence of the Slater point (x^s,u^s,t^s).

[8] Note that F_1 is proper and closed thanks to the existence of the Slater point (x^s,u^s,t^s).

at $U(\bar{x},\bar{u},f(\bar{x}))$ with exponent $\alpha\in[0,1)$, where U is the same as in (i).[9] *Then F satisfies the KL property at $(\bar{x},\bar{u},f(\bar{x}))$ with exponent α.*

Proof Since $0\in\partial F(\bar{x},\bar{u},f(\bar{x}))$, we have in view of [46, Exercise 8.8] that

$$0_{n+N+1}\in(0_n,0_N,1)+N_{\mathfrak{D}}(\bar{x},\bar{u},f(\bar{x})), \tag{4.3}$$

where $\mathfrak{D}$ is defined as in (3.7), and 0_k is the zero vector of dimension k. Next, since $\delta_{\mathfrak{D}}(x,u,t)=[\delta_{\mathcal{S}^d_+-A_{00}}\circ\mathcal{A}^*](x,u,t)$ and we have $\mathcal{A}^*(x^s,u^s,t^s)\succ -A_{00}$ by assumption, using [45, Theorem 23.9], we deduce that

$$N_{\mathfrak{D}}(\bar{x},\bar{u},f(\bar{x}))=\partial\Big[\delta_{\mathcal{S}^d_+-A_{00}}\circ\mathcal{A}^*\Big](\bar{x},\bar{u},f(\bar{x}))=\mathcal{A}N_{\mathcal{S}^d_+-A_{00}}(\mathcal{A}^*(\bar{x},\bar{u},f(\bar{x}))).$$

This together with (4.3) implies that there exists $Y\in N_{\mathcal{S}^d_+-A_{00}}(\mathcal{A}^*(\bar{x},\bar{u},f(\bar{x})))$ such that

$$\langle A_1,Y\rangle=\cdots=\langle A_n,Y\rangle=\langle B_1,Y\rangle=\cdots=\langle B_N,Y\rangle=0 \;\text{ but }\; \langle A_0,Y\rangle=-1;$$

in particular, $A_0\notin\operatorname{span}\{A_1,\ldots,A_n,B_1,\ldots,B_N\}$ and hence $A_0\neq 0$.

If $\operatorname{span}\{A_1,\ldots,A_n,B_1,\ldots,B_N\}=\{0\}$, then $A_i=B_j=0$ for $i=1,\ldots,n$ and $j=1,\ldots,N$. In this case, set $\hat{A}_0=A_0$. We see that $\{\hat{A}_0\}$ is an orthogonal set and we have

$$\begin{bmatrix}\mathbf{a}_1 \;\ldots\; \mathbf{a}_n \;\mathbf{b}_1 \;\ldots\; \mathbf{b}_N \;\mathbf{a}_0\end{bmatrix}=\hat{\mathbf{a}}_0\begin{bmatrix}0^T_{n+N}\;1\end{bmatrix},$$

where 0_{n+N} is the zero vector of dimension $n+N$. Thus, the conclusion in (i) holds in this case.

Otherwise, $\operatorname{span}\{A_1,\ldots,A_n,B_1,\ldots,B_N\}\neq\{0\}$ and we let $\{\bar{A}_1,\ldots,\bar{A}_p\}$ be a maximal linearly independent subset of $\{A_1,\ldots,A_n,B_1,\ldots,B_N\}$. Then there exists $M_0\in\mathbb{R}^{p\times(n+N)}$ with full row rank such that $[\mathbf{a}_1\;\ldots\;\mathbf{a}_n\;\mathbf{b}_1\;\ldots\;\mathbf{b}_N]=[\bar{\mathbf{a}}_1\;\ldots\;\bar{\mathbf{a}}_p]\,M_0$, where $\bar{\mathbf{a}}_i\in\mathbb{R}^{d^2}$ is the columnwise vectorization of $\bar{A}_i$. Thus

$$[\mathbf{a}_1\;\ldots\;\mathbf{a}_n\;\mathbf{b}_1\;\ldots\;\mathbf{b}_N\;\mathbf{a}_0]=[\bar{\mathbf{a}}_1\;\ldots\;\bar{\mathbf{a}}_p\;\mathbf{a}_0]\begin{bmatrix}M_0 & 0\\ 0 & 1\end{bmatrix}. \tag{4.4}$$

Using Gram-Schmidt process followed by a suitable scaling to $\{\bar{A}_1,\ldots,\bar{A}_p,A_0\}$, there exists an invertible upper triangle matrix $U_0\in\mathbb{R}^{(p+1)\times(p+1)}$ with the $(U_0)_{p+1,p+1}=1$ and an orthogonal basis $\{\hat{A}_1,\ldots,\hat{A}_p,\hat{A}_0\}$ of $\operatorname{span}\{\bar{A}_1,\ldots,\bar{A}_p,A_0\}$ such that $[\bar{\mathbf{a}}_1\;\ldots\;\bar{\mathbf{a}}_p\;\mathbf{a}_0]=\begin{bmatrix}\hat{\mathbf{a}}_1\;\ldots\;\hat{\mathbf{a}}_p\;\hat{\mathbf{a}}_0\end{bmatrix}U_0$, where $\hat{\mathbf{a}}_i$ is the columnwise vectorization of $\hat{A}_i$. This together with (4.4) shows that

$$\begin{bmatrix}\mathbf{a}_1\;\ldots\;\mathbf{a}_n\;\mathbf{b}_1\;\ldots\;\mathbf{b}_N\;\mathbf{a}_0\end{bmatrix}=[\bar{\mathbf{a}}_1\;\ldots\;\bar{\mathbf{a}}_p\;\mathbf{a}_0]\begin{bmatrix}M_0 & 0\\ 0 & 1\end{bmatrix}=\begin{bmatrix}\hat{\mathbf{a}}_1\;\ldots\;\hat{\mathbf{a}}_p\;\hat{\mathbf{a}}_0\end{bmatrix}U,$$

[9] Here and henceforth, $U(\bar{x},\bar{u},f(\bar{x}))$ is a short-hand notation for the matrix vector product $U\begin{bmatrix}\bar{x}\\ \bar{u}\\ f(\bar{x})\end{bmatrix}$.

where $U := U_0 \begin{bmatrix} M_0 & 0 \\ 0 & 1 \end{bmatrix}$ has full row rank and the entries of the $(p+1)^{\text{th}}$ row of U are 0 except for $U_{p+1,n+N+1} = 1$. This proves (i).

Now, using the definition of F_1 in (4.2), we have $F(x,u,t) = F_1(U(x,u,t))$. Since U is surjective and the KL exponent of F_1 is α at $U(\bar{x},\bar{u},f(\bar{x}))$, using a similar argument as in [34, Theorem 3.2], the KL exponent of F at $(\bar{x},\bar{u},f(\bar{x}))$ equals α. This completes the proof. $\square$

Finally, we rewrite F_1 in (4.2) suitably as a function on $\mathcal{S}^d$ that satisfies a certain "strict complementarity" condition so that Lemma 2.3 can be readily applied to deducing the KL exponent of F_1 explicitly.

Lemma 4.3 *Let f be a proper closed function and $\bar{x} \in \operatorname{dom} f$. Suppose in addition that f is SDP-representable with its epigraph represented as in* (3.5). *Let F be defined as in* (3.7)*, $\mathfrak{D}_{\bar{x}}$ be defined as in* (4.1)*, and $\bar{u} \in \mathfrak{D}_{\bar{x}}$. Suppose that the following conditions hold:*

(i) **(Slater's condition)** *There exists (x^s,u^s,t^s) such that $A_{00}+\mathcal{A}^*(x^s,u^s,t^s) \succ 0$, where A_{00} and $\mathcal{A}$ are given in* (3.5) *and* (3.6) *respectively.*[10]
(ii) **(Strict complementarity)** *It holds that* $0 \in \operatorname{ri} \partial F(\bar{x},\bar{u},f(\bar{x}))$.

Let F_1 be defined as in (4.2)*. Then $U(\bar{x},\bar{u},f(\bar{x})) \in \operatorname{dom} \partial F_1$ and F_1 satisfies the KL property at $U(\bar{x},\bar{u},f(\bar{x}))$ with exponent $\frac{1}{2}$, where U is given in Lemma 4.2*(i).

Proof Define $\bar{\mathcal{A}} : \mathcal{S}^d \to \mathbb{R}^{p+1}$ by

$$\bar{\mathcal{A}}(W) := \left[\langle \hat{A}_1, W\rangle \ \ldots \ \langle \hat{A}_p, W\rangle \ \langle \hat{A}_0, W\rangle\right]^T,$$

where $\{\hat{A}_0,\ldots,\hat{A}_p\}$ is given by Lemma 4.2(i). Since $\{\hat{A}_0,\ldots,\hat{A}_p\}$ is orthogonal, we see that $\bar{\mathcal{A}}$ is surjective and $\bar{\mathcal{A}}^* : \mathbb{R}^{p+1} \to \mathcal{S}^d$ with $\bar{\mathcal{A}}^*(z,t) := \hat{A}_0 t + \sum_{w=1}^p \hat{A}_w z_w$ is injective. Also, for any $(z,t) \in \mathbb{R}^{p+1}$, by orthogonality,

$$\bar{\mathcal{A}}\bar{\mathcal{A}}^*(z,t) = \bar{\mathcal{A}}\left(\hat{A}_0 t + \sum_{w=1}^p \hat{A}_w z_w\right) = \left(\|\hat{A}_1\|_F^2 z_1,\ldots,\|\hat{A}_p\|_F^2 z_p, \|\hat{A}_0\|_F^2 t\right).$$

Choose a basis $\{H_1,H_2,\ldots,H_r\}$ of $\ker \bar{\mathcal{A}}$ and define a linear map $\mathcal{H} : \mathcal{S}^d \to \mathbb{R}^r$ by[11]

$$\mathcal{H}(W) := \left[\langle H_1, W\rangle \ \cdots \ \langle H_r, W\rangle\right]^T. \tag{4.5}$$

Define a proper closed function $F_2 : \mathcal{S}^d \to \mathbb{R} \cup \{\infty\}$ by

$$\begin{aligned} &F_2(X) := \|\hat{A}_0\|_F^{-2}\langle \hat{A}_0, X\rangle + \delta_{\mathfrak{D}_2}(X) \\ &\text{with } \mathfrak{D}_2 := \left\{X \in \mathcal{S}_+^d : \ \mathcal{H}X = \mathcal{H}A_{00}\right\}. \end{aligned} \tag{4.6}$$

[10] Note that this condition implies that both F in (3.7) and F_1 in (4.2) are proper and closed.

[11] In the case when $\ker \bar{\mathcal{A}} = \{0\}$ so that the basis is empty (i.e., $r = 0$), we define $\mathcal{H}$ to be the *unique* linear map that maps $\mathcal{S}^d$ onto the *zero vector space*.

Thanks to the identity $(\ker \bar{\mathcal{A}})^\perp = \mathrm{Range}\,(\bar{\mathcal{A}}^*)$ and the fact that $\mathcal{H}X = \mathcal{H}A_{00}$ if and only if $X - A_{00} \in (\ker \bar{\mathcal{A}})^\perp$, we have the following relations concerning $\mathfrak{D}_2$ and the $\mathfrak{D}_1$ defined in (4.2):

$$\begin{gathered}(z,t) \in \mathfrak{D}_1 \Longrightarrow A_{00} + \bar{\mathcal{A}}^*(z,t) \in \mathfrak{D}_2,\\ X \in \mathfrak{D}_2 \Rightarrow \exists \text{ unique } (z,t) \text{ s.t. } A_{00} + \bar{\mathcal{A}}^*(z,t) = X, \text{ and } (z,t) \in \mathfrak{D}_1,\end{gathered} \tag{4.7}$$

where the second implication also makes use of the injectivity of $\bar{\mathcal{A}}^*$. We then deduce further that for any $(z,t) \in \mathbb{R}^{p+1}$,

$$\begin{aligned}&F_2(A_{00} + \bar{\mathcal{A}}^*(z,t)) - \|\hat{A}_0\|_F^{-2}\langle \hat{A}_0, A_{00}\rangle\\ &= \langle \bar{\mathcal{A}}\left(\|\hat{A}_0\|_F^{-2}\hat{A}_0\right), (z,t)\rangle + \delta_{\mathfrak{D}_2}(A_{00} + \bar{\mathcal{A}}^*(z,t))\\ &= t + \delta_{\mathfrak{D}_2}(A_{00} + \bar{\mathcal{A}}^*(z,t)) = F_1(z,t),\end{aligned} \tag{4.8}$$

where the last equality follows from (4.7).

Next, let U be as in Lemma 4.2(i). Since the entries in the $(p+1)^{\text{th}}$ row of U are 0 except for $U_{p+1,n+N+1} = 1$, there exists $\bar{z} \in \mathbb{R}^p$ such that[12]

$$U(\bar{x}, \bar{u}, f(\bar{x})) = (\bar{z}, f(\bar{x})). \tag{4.9}$$

Now, define

$$\bar{X} := A_{00} + \bar{\mathcal{A}}^*(\bar{z}, f(\bar{x})). \tag{4.10}$$

We claim that $0 \in \mathrm{ri}\,\partial F_2(\bar{X})$. We first show that

$$0 \in \mathrm{ri}\,\partial F_1(\bar{z}, f(\bar{x})). \tag{4.11}$$

In fact, using [45, Theorem 23.9] (note that $U(x^s, u^s, t^s) \in \mathrm{int}\,\mathfrak{D}_1$ thanks to assumption (i)) together with the assumption (ii), we have

$$0 \in \mathrm{ri}\,\partial F(\bar{x}, \bar{u}, f(\bar{x})) = \mathrm{ri}\,\left[U^T \partial F_1\left(U(\bar{x}, \bar{u}, f(\bar{x}))\right)\right] = U^T \mathrm{ri}\,\partial F_1\left(U(\bar{x}, \bar{u}, f(\bar{x}))\right),$$

where the second equality follows from [45, Theorem 6.6]. Since U has full row rank and thus U^T is injective, recalling the definition of $\bar{z}$ in (4.9), we deduce further that (4.11) holds. Now, using this and [46, Exercise 8.8], we have

$$0 \in \mathrm{ri}\,\partial F_1(\bar{z}, f(\bar{x})) = (\underbrace{0, \ldots, 0}_{p \text{ entries}}, 1) + \mathrm{ri}\,N_{\mathfrak{D}_1}(\bar{z}, f(\bar{x})). \tag{4.12}$$

Now, notice that $\delta_{\mathfrak{D}_1}(z,t) = \left[\delta_{S_+^d - A_{00}} \circ \bar{\mathcal{A}}^*\right](z,t)$ and

$$\mathfrak{D}_2 \ni X^s := A_{00} + \bar{\mathcal{A}}^*(z^s, t^s) = A_{00} + \mathcal{A}^*(x^s, u^s, t^s) \succ 0 \tag{4.13}$$

with $(z^s, t^s) = U(x^s, u^s, t^s)$, where the inclusion holds thanks to (4.7). Using these and [45, Theorem 23.9], we see that

$$\mathrm{ri}\,N_{\mathfrak{D}_1}(\bar{z}, f(\bar{x})) = \mathrm{ri}\,\partial\left[\delta_{S_+^d - A_{00}} \circ \bar{\mathcal{A}}^*\right](\bar{z}, f(\bar{x})) = \mathrm{ri}\,\bar{\mathcal{A}} N_{S_+^d}(\bar{X}) = \bar{\mathcal{A}}\,\mathrm{ri}\,N_{S_+^d}(\bar{X}),$$

[12] Recall that $p \geq 0$. When $p = 0$, we interpret $\bar{z}$ as a null vector so that $U(\bar{x}, \bar{u}, f(\bar{x})) = f(\bar{x})$.

where the last equality follows from [45, Theorem 6.6]. This together with (4.12) implies that there exists $\tilde{Y} \in \operatorname{ri} N_{\mathcal{S}^d_+}(\bar{X})$ such that

$$\langle \hat{A}_1, \tilde{Y} \rangle = \cdots = \langle \hat{A}_p, \tilde{Y} \rangle = 0 \text{ and } \langle \hat{A}_0, \tilde{Y} \rangle = -1. \tag{4.14}$$

The second relation in (4.14) gives $\langle \hat{A}_0, \tilde{Y} + \|\hat{A}_0\|_F^{-2} \hat{A}_0 \rangle = \langle \hat{A}_0, \tilde{Y} \rangle + 1 = 0$. In addition, in view of the first relation in (4.14) and the orthogonality of $\{\hat{A}_0, \ldots, \hat{A}_p\}$, we have $\langle \hat{A}_i, \tilde{Y} + \|\hat{A}_0\|_F^{-2} \hat{A}_0 \rangle = \langle \hat{A}_i, \tilde{Y} \rangle + \langle \hat{A}_i, \|\hat{A}_0\|_F^{-2} \hat{A}_0 \rangle = 0$ for all $i = 1, \ldots, p$. Thus, it holds that $\tilde{Y} + \|\hat{A}_0\|_F^{-2} \hat{A}_0 \in \ker \bar{\mathcal{A}}$. Hence, there exists $\omega \in \mathbb{R}^r$ such that

$$\tilde{Y} + \|\hat{A}_0\|_F^{-2} \hat{A}_0 = \sum_{i=1}^{r} H_i \omega_i \tag{4.15}$$

with r and H_i defined as in (4.5).[13] Using (4.15) and the definition of $\tilde{Y}$, we have further that

$$0 = \tilde{Y} + \|\hat{A}_0\|_F^{-2} \hat{A}_0 - \sum_{i=1}^{r} H_i \omega_i \in \operatorname{ri} N_{\mathcal{S}^d_+}(\bar{X}) + \|\hat{A}_0\|_F^{-2} \hat{A}_0 + \operatorname{Range} \mathcal{H}^*. \tag{4.16}$$

On the other hand, using the definition of F_2 in (4.6), we have

$$\begin{aligned}
\operatorname{ri} \partial F_2(\bar{X}) &= \|\hat{A}_0\|_F^{-2} \hat{A}_0 + \operatorname{ri} \partial \delta_{\mathfrak{D}_2}(\bar{X}) \\
&= \|\hat{A}_0\|_F^{-2} \hat{A}_0 + \operatorname{ri} \left(N_{\mathcal{H}^{-1}\{\mathcal{H}A_{00}\}}(\bar{X}) + N_{\mathcal{S}^d_+}(\bar{X}) \right) \\
&= \|\hat{A}_0\|_F^{-2} \hat{A}_0 + \operatorname{ri} N_{\mathcal{H}^{-1}\{\mathcal{H}A_{00}\}}(\bar{X}) + \operatorname{ri} N_{\mathcal{S}^d_+}(\bar{X}) \\
&= \|\hat{A}_0\|_F^{-2} \hat{A}_0 + \operatorname{Range} \mathcal{H}^* + \operatorname{ri} N_{\mathcal{S}^d_+}(\bar{X}),
\end{aligned}$$

where the second equality follows from [45, Theorem 23.8] and (4.13), and the third equality follows from [45, Corollary 6.6.2]. This together with (4.16) shows

$$0 \in \operatorname{ri} \partial F_2(\bar{X}). \tag{4.17}$$

In view of (4.13) and (4.17), we can now apply Lemma 2.3 and deduce that, for a given compact neighborhood $\mathfrak{U}$ of $\bar{X}$, there exists $c > 0$ such that for any $X \in \mathfrak{U} \cap \mathfrak{D}_2$,

$$\operatorname{dist}(X, \operatorname{Arg\,min} F_2) \leq c \left(F_2(X) - F_2(\bar{X}) \right)^{\frac{1}{2}}. \tag{4.18}$$

Thus, fix an $\epsilon > 0$ so that $A_{00} + \bar{\mathcal{A}}^*(z,t) \in \mathfrak{U}$ whenever $(z,t) \in B((\bar{z}, f(\bar{x})), \epsilon)$; such an ϵ exists thanks to the definitions of $\bar{z}$ in (4.9) and $\bar{X}$ in (4.10). Now, consider any (z,t) satisfying $(z,t) \in B((\bar{z}, f(\bar{x})), \epsilon)$ and $F_1(\bar{z}, f(\bar{x})) < F_1(z,t) < F_1(\bar{z}, f(\bar{x})) + \epsilon$. Then $(z,t) \in \operatorname{dom} F_1$, which means $A_{00} + \bar{\mathcal{A}}^*(z,t) \in \mathfrak{D}_2$ according to (4.7). Hence, using (4.18), we have

$$\operatorname{dist}^2((z,t), \operatorname{Arg\,min} F_1) \leq \|(z,t) - (z^*, t^*)\|^2 \overset{\text{(a)}}{\leq} c_1 \left\| \bar{\mathcal{A}}^*(z,t) - \bar{\mathcal{A}}^*(z^*, t^*) \right\|_F^2$$

[13] In the case when $\ker \bar{\mathcal{A}} = \{0\}$ (i.e., $r = 0$), we have $\tilde{Y} + \|\hat{A}_0\|_F^{-2} \hat{A}_0 = 0$. In this case, we interpret ω as a null vector.

$$= c_1\|A_{00} + \bar{\mathcal{A}}^*(z,t) - X^*\|_F^2 = c_1\mathrm{dist}^2(A_{00} + \bar{\mathcal{A}}^*(z,t), \mathrm{Arg\,min}\, F_2)$$
$$\le c^2 c_1 \left(F_2(A_{00} + \bar{\mathcal{A}}^*(z,t)) - F_2(\bar{X})\right) \overset{\text{(b)}}{=} c^2 c_1 \left(F_1(z,t) - F_1(\bar{z}, f(\bar{x}))\right),$$

where X^* denotes the projection of $A_{00} + \bar{\mathcal{A}}^*(z,t)$ on $\mathrm{Arg\,min}\, F_2$ and (z^*,t^*) is the corresponding element in $\mathrm{Arg\,min}\, F_1$ such that $X^* = A_{00} + \bar{\mathcal{A}}^*(z^*,t^*)$ (the existence of (z^*,t^*) follows from (4.7) and (4.8)), (a) holds for some $c_1 > 0$ because $\bar{\mathcal{A}}^*$ is injective, and (b) follows from (4.8). Combining this with [14, Theorem 5], we conclude that F_1 satisfies the KL property with exponent $\frac{1}{2}$ at $(\bar{z}, f(\bar{x})) = U(\bar{x}, \bar{u}, f(\bar{x}))$. □

We are now ready to state and prove our main result in this section.

Theorem 4.1 (KL exponent of SDP-representable functions) *Let f be a proper closed function and $\bar{x} \in \mathrm{dom}\,\partial f$. Suppose in addition that f is SDP-representable with its epigraph represented as in* (3.5) *and that the following conditions hold:*

- (i) **(Slater's condition)** *There exists (x^s, u^s, t^s) such that $A_{00} + \mathcal{A}^*(x^s, u^s, t^s) \succ 0$, where A_{00} and $\mathcal{A}$ are given in* (3.5) *and* (3.6) *respectively.*
- (ii) **(Compactness)** *The set $\mathfrak{D}_{\bar{x}}$ defined as in* (4.1) *is nonempty and compact.*
- (iii) **(Strict complementarity)** *It holds that $0 \in \mathrm{ri}\,\partial F(\bar{x}, u, f(\bar{x}))$ for all $u \in \mathfrak{D}_{\bar{x}}$, where F is defined as in* (3.7) *and $\mathfrak{D}_{\bar{x}}$ is defined as in* (4.1).[14]

Then f satisfies the KL property at $\bar{x}$ with exponent $\frac{1}{2}$.

Remark 4.1 In Theorem 4.1, we require $0 \in \mathrm{ri}\,\partial F(\bar{x}, u, f(\bar{x}))$ for all $u \in \mathfrak{D}_{\bar{x}}$ with $\mathfrak{D}_{\bar{x}}$ defined as in (4.1). This can be hard to check in practice. In Sections 4.2 and 4.3, we will impose additional assumptions on f so that this condition can be replaced by $0 \in \mathrm{ri}\,\partial f(\bar{x})$, which is a form of strict complementarity condition imposed on the original function f (rather than the representation F in the lifted space).

Proof In view of Lemma 4.1, it suffices to show that F satisfies the KL property with exponent $\frac{1}{2}$ at every point in $\{\bar{x}\} \times \mathfrak{D}_{\bar{x}} \times \{f(\bar{x})\}$. Fix any $\bar{u} \in \mathfrak{D}_{\bar{x}}$. From Lemma 4.3, we know that F_1 defined as in (4.2) has KL property with exponent $\frac{1}{2}$ at $U(\bar{x}, \bar{u}, f(\bar{x})) \in \mathrm{dom}\,\partial F_1$, where U is given in Lemma 4.2(i). Using this together with Lemma 4.2, we know that F satisfies the KL property with exponent $\frac{1}{2}$ at $(\bar{x}, \bar{u}, f(\bar{x}))$. This completes the proof. □

We would like to point out that the third condition in Theorem 4.1 cannot be replaced by "$0 \in \mathrm{ri}\,\partial f(\bar{x})$" in general. One concrete counter-example is $f(x) = x^4$. Indeed, for this function, the global minimizer is 0 and we have

[14] We note that because of the Slater's condition, the function F in (3.7) is proper and closed.

$\partial f(0) = \{\nabla f(0)\} = \{0\}$, which implies that $0 \in \operatorname{ri} \partial f(0)$. Moreover, this function is SDP-representable:

$$\operatorname{epi} f = \left\{ (x,t) \in \mathbb{R}^n \times \mathbb{R} : \begin{bmatrix} 1 & y & 0 & 0 \\ y & t & 0 & 0 \\ 0 & 0 & 1 & x \\ 0 & 0 & x & y \end{bmatrix} \succeq 0 \text{ for some } y \right\}.$$

It is easy to check that the first two conditions of Theorem 4.1 are satisfied for $\bar{x} = 0$. However, it can be directly verified that this f does not have KL property with exponent $\frac{1}{2}$ at 0. This concrete example suggests that the third condition in Theorem 4.1 cannot be replaced by $0 \in \operatorname{ri} \partial f(\bar{x})$ in general.

Next, in Sections 4.2 and 4.3, we will look at special SDP-representable functions and show that the third condition in Theorem 4.1 can indeed be replaced by $0 \in \operatorname{ri} \partial f(\bar{x})$ in those cases.

4.2 Sum of LMI-representable functions

In this section, we discuss how the KL exponent of the sum of finitely many proper closed LMI-representable functions as defined in Section 3.1.1(ii) can be deduced through Theorem 4.1. Compared with Theorem 4.1, the strict complementarity condition in this section is now imposed *directly* on the original function.

Theorem 4.2 (KL exponent of sum of LMI-representable functions) *Let $f = \sum_{i=1}^m f_i$, where each $f_i : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is proper and closed. Suppose that each f_i is LMI-representable, i.e., there exist $d_i \geq 1$ and matrices $\{A_{00}^i, A_0^i, A_1^i, \ldots, A_n^i\} \subset \mathcal{S}^{d_i}$ such that*

$$\operatorname{epi} f_i = \left\{ (x,t) \in \mathbb{R}^n \times \mathbb{R} : \; A_{00}^i + \sum_{j=1}^n A_j^i x_j + A_0^i t \succeq 0 \right\}.$$

Suppose in addition that there exist $x^s \in \mathbb{R}^n$ and $s^s \in \mathbb{R}^m$ such that for $i = 1, \ldots, m$,

$$A_{00}^i + \sum_{j=1}^n A_j^i x_j^s + A_0^i s_i^s \succ 0.$$

If $\bar{x} \in \operatorname{dom} \partial f$ satisfies $0 \in \operatorname{ri} \partial f(\bar{x})$, then f satisfies the KL property at $\bar{x}$ with exponent $\frac{1}{2}$.

Proof We first derive an SDP representation of $\operatorname{epi} f$. To this end, define

$$\hat{\mathfrak{D}} := \left\{ (x,s,t) : \; t \geq \sum_{i=1}^m s_i \text{ and } s_i \geq f_i(x), \; \forall i = 1, \ldots, m \right\}.$$

Then it holds that $(x, s, t) \in \hat{\mathfrak{D}}$ if and only if

$$\begin{bmatrix} t - \sum_{i=1}^{m} s_i & 0 & \cdots & 0 \\ 0 & A_{00}^1 + \sum_{j=1}^{n} A_j^1 x_j + A_0^1 s_1 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & A_{00}^m + \sum_{j=1}^{n} A_j^m x_j + A_0^m s_m \end{bmatrix} \succeq 0. \tag{4.19}$$

Since

$$(x, t) \in \operatorname{epi} f \iff t \geq \sum_{i=1}^{m} f_i(x) \iff \exists s \in \mathbb{R}^m \text{ s.t. } (x, s, t) \in \hat{\mathfrak{D}}, \tag{4.20}$$

we see that f is SDP-representable. Moreover, if we define

$$F(x, s, t) := t + \delta_{\hat{\mathfrak{D}}}(x, s, t), \tag{4.21}$$

then it holds that $f(x) = \inf_{s,t} F(x, s, t)$ for all $x \in \mathbb{R}^n$. We next show that f and the F defined in (4.21) satisfy the conditions required in Theorem 4.1.

First, from the definition of $x^s \in \mathbb{R}^n$ and $s^s \in \mathbb{R}^m$, we have

$$\begin{bmatrix} t^s - \sum_{i=1}^{m} s_i^s & 0 & \cdots & 0 \\ 0 & A_{00}^1 + \sum_{j=1}^{n} A_j^1 x_j^s + A_0^1 s_1^s & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & A_{00}^m + \sum_{j=1}^{n} A_j^m x_j^s + A_0^m s_m^s \end{bmatrix} \succ 0,$$

where $t^s := \sum_{i=1}^{m} s_i^s + 1$. This together with (4.19) and (4.20) shows that condition (i) in Theorem 4.1 holds.

Next, note that the set $\{s :\ (\bar{x}, s, f(\bar{x})) \in \hat{\mathfrak{D}}\} = \{(f_1(\bar{x}), \ldots, f_m(\bar{x}))\}$, which is clearly nonempty and compact. In view of this and (4.21), we conclude that condition (ii) in Theorem 4.1 is satisfied.

Finally, we look at the strict complementarity condition, i.e., condition (iii) in Theorem 4.1. Notice that the definition of $x^s \in \mathbb{R}^n$ implies

$$x^s \in \bigcap_{i=1}^{m} \operatorname{int} \operatorname{dom} f_i. \tag{4.22}$$

Write $\bar{s} := (f_1(\bar{x}), \cdots, f_m(\bar{x}))$ for notational simplicity. Define

$$\mathfrak{C}_0 = \left\{ (x, s, t) :\ t \geq \sum_{i=1}^{m} s_i \right\} \text{ and } \mathfrak{C}_i = \{(x, s, t) :\ s_i \geq f_i(x)\},\ \forall i = 1, \ldots, m.$$

Then $\hat{\mathfrak{D}} = \bigcap_{i=0}^{m} \mathfrak{C}_i$. Moreover, using [45, Theorem 7.6], we have for $i = 1, \ldots, m$ that

$$\begin{aligned} \operatorname{ri} \mathfrak{C}_i &= \operatorname{ri} \{(x, s, t) :\ g_i(x, s, t) \leq 0\} = \{(x, s, t) \in \operatorname{ri} \operatorname{dom} g_i :\ g_i(x, s, t) < 0\} \\ &= \{(x, s, t) \in \operatorname{ri} \operatorname{dom} f_i \times \mathbb{R}^m \times \mathbb{R} :\ g_i(x, s, t) < 0\}, \end{aligned}$$

where $g_i(x,s,t) = f_i(x) - s_i$ for each i. This together with (4.22) shows that $\bigcap_{i=0}^m \mathrm{ri}\, \mathfrak{C}_i \neq \emptyset$. Using this, [45, Theorem 23.8] and the definition of F in (4.21), we have

$$\partial F(\bar{x}, \bar{s}, f(\bar{x})) = (0_{n+m}, 1) + \sum_{i=0}^{m} N_{\mathfrak{C}_i}(\bar{x}, \bar{s}, f(\bar{x})), \tag{4.23}$$

where 0_p is the zero vector of dimension p, and recall that $\bar{s} = (f_1(\bar{x}), \cdots, f_m(\bar{x}))$.

We claim that $0 \in \mathrm{ri}\, \partial F(\bar{x}, \bar{s}, f(\bar{x}))$. To this end, note first that the assumption $0 \in \mathrm{ri}\, \partial f(\bar{x})$ and (4.22) together with [45, Theorem 23.8] imply that $\bar{x} \in \bigcap_i \mathrm{dom}\, \partial f_i$. Hence, we have from [45, Theorem 23.7] that for each $i = 1, \ldots, m$,

$$\begin{aligned} N_{\mathfrak{C}_i}(\bar{x}, \bar{s}, f(\bar{x})) &= \mathrm{cl}\ [\mathrm{cone}\, \partial g_i(\bar{x}, \bar{s}, f(\bar{x}))] \\ &= \mathrm{cl} \bigcup_{\lambda_i \geq 0} (\lambda_i \partial f_i(\bar{x}), 0_{i-1}, -\lambda_i, 0_{m+1-i}) \end{aligned} \tag{4.24}$$

where the second equality follows from [46, Proposition 10.5] and $\mathrm{cone}\, \mathfrak{B}$ denotes the convex conical hull of $\mathfrak{B}$. Similarly, we also have

$$N_{\mathfrak{C}_0}(\bar{x}, \bar{s}, f(\bar{x})) = \mathrm{cl} \bigcup_{\lambda_0 \geq 0} (0_n, \lambda_0 \cdot 1_m, -\lambda_0)\,, \tag{4.25}$$

where 1_m is the m-dimensional vector of all ones. Using (4.23), (4.24) and (4.25), we have

$$\begin{aligned} &- (0_{n+m}, 1) + \mathrm{ri}\, \partial F(\bar{x}, \bar{s}, f(\bar{x})) \overset{\text{(a)}}{=} \sum_{i=0}^{m} \mathrm{ri}\, N_{\mathfrak{C}_i}(\bar{x}, \bar{s}, f(\bar{x})) \\ &\overset{\text{(b)}}{=} \sum_{i=1}^{m} \mathrm{ri} \left[\mathrm{cl} \bigcup_{\lambda_i \geq 0} (\lambda_i \partial f_i(\bar{x}), 0_{i-1}, -\lambda_i, 0_{m+1-i}) \right] + \mathrm{ri} \left[\mathrm{cl} \bigcup_{\lambda_0 \geq 0} (0_n, \lambda_0 \cdot 1_m, -\lambda_0) \right] \\ &\overset{\text{(c)}}{=} \sum_{i=1}^{m} \bigcup_{\lambda_i > 0} (\lambda_i\, \mathrm{ri}\, \partial f_i(\bar{x}), 0_{i-1}, -\lambda_i, 0_{m+1-i}) + \bigcup_{\lambda_0 > 0} (0_n, \lambda_0 \cdot 1_m, -\lambda_0) \end{aligned}$$

where (a) follows from (4.23) and [45, Corollary 6.6.2], (b) follows from (4.24) and (4.25), and (c) follows from [45, Theorem 6.3] and [45, Corollary 6.8.1]. This together with $0 \in \mathrm{ri}\, \partial f(\bar{x})$ yields

$$\begin{aligned} 0 \in (\mathrm{ri}\, \partial f(\bar{x}), 0_m, 0) &= (0_n, 0_m, 1) + (\mathrm{ri}\, \partial f(\bar{x}), -1_m, 0) + (0_n, 1_m, -1) \\ &= (0_n, 0_m, 1) + \left(\sum_{i=1}^{m} \mathrm{ri}\, \partial f_i(\bar{x}), -1_m, 0 \right) + (0_n, 1_m, -1) \subseteq \mathrm{ri}\, \partial F(\bar{x}, \bar{s}, f(\bar{x})), \end{aligned}$$

where the second equality follows from [45, Theorem 23.8] and [45, Corollary 6.6.2], thanks to (4.22). Thus, condition (iii) in Theorem 4.1 is also satisfied. The desired conclusion now follows from Theorem 4.1. □

Example 4.1 Note that ℓ_1-norm, ℓ_2-norm, convex quadratic functions and indicator functions of second-order cones are all LMI-representable. Using these, we can infer from Theorem 4.2 that the following functions f satisfy the KL property with exponent $\frac{1}{2}$ at any $\bar{x}$ that verifies $0 \in \operatorname{ri} \partial f(\bar{x})$:

(i) **Group Lasso with overlapping blocks of variables:**

$$f(x) = \frac{1}{2}\|Ax - b\|^2 + \sum_{i=1}^{s} w_i \|x_{J_i}\|,$$

where $b \in \mathbb{R}^p$, $A \in \mathbb{R}^{p\times n}$, $J_i \subseteq \{1, \ldots, n\}$ with $\bigcup_{i=1}^{s} J_i = \{1, \ldots, n\}$, x_{J_i} is the subvector of x indexed by J_i, and $w_i \geq 0$, $i = 1, \ldots, s$. We emphasize here that $J_i \cap J_j$ can be nonempty when $i \neq j$.

(ii) **Least squares with products of second-order cone constraints:**

$$f(x) = \frac{1}{2}\|Ax - b\|^2 + \delta_{\prod_{i=1}^{s} \mathrm{SOC}_{n_i}}(x),$$

where $b \in \mathbb{R}^p$, $A \in \mathbb{R}^{p\times n}$, $x = (x_1, \ldots, x_s) \in \prod_{i=1}^{s} \mathbb{R}^{n_i}$ with $x_i \in \mathbb{R}^{n_i}$, $i = 1, \ldots, s$, and SOC_{n_i} is the second-order cone in $\mathbb{R}^{n_i}$.

(iii) **Group fused Lasso [2]:**

$$f(x) = \frac{1}{2}\|Ax - b\|^2 + \sum_{i=1}^{s} w_i \|x_{J_i}\| + \sum_{i=2}^{s} \nu_i \|x_{J_i} - x_{J_{i-1}}\|,$$

where $b \in \mathbb{R}^p$, $A \in \mathbb{R}^{p\times n}$ with $n = rs$ for some $r \in \mathbb{N}$, J_i is an equipartition of $\{1, \ldots, n\}$ in the sense that $\bigcup_{i=1}^{s} J_i = \{1, \ldots, n\}$, $J_i \cap J_j = \emptyset$ and $|J_i| = |J_j| = r$ for $i \neq j$, w_i, $\nu_i \geq 0$, $i = 1, \ldots, s$.

4.3 Sum of LMI-representable functions and the nuclear norm

In this section, we apply Theorem 4.2 and Corollary 3.3 to derive the KL exponent of the function in (3.8) under suitable assumptions. It is known (see, for example [44]) that the nuclear norm can be expressed as

$$\|X\|_* = \frac{1}{2} \inf_{U,V} \left\{ \mathrm{tr}(U) + \mathrm{tr}(V) : \begin{bmatrix} U & X \\ X^T & V \end{bmatrix} \succeq 0,\, U \in \mathcal{S}^m, V \in \mathcal{S}^n \right\} \tag{4.26}$$

for any $X \in \mathbb{R}^{m\times n}$. This fact plays an important role for our analysis later on, and shows that the nuclear norm is an SDP representable function. To the best of our knowledge, it is not known that whether the nuclear norm is LMI representable. Our analysis is an attempt to generalize our results on the sum of LMI representable functions (with strict complementarity assumption on the original function) to a large subclass of SDP representable functions that arises in many important areas such as matrix completion [44].

Theorem 4.3 (KL exponent of sum of LMI-representable functions and the nuclear norm) *Let f be defined as in* (3.8) *and let symmetric matrices A_{00}^k, A_0^k, A_{ij}^k, $i=1,\ldots,m$ and $j=1,\ldots,n$, be such that*

$$\operatorname{epi} f_k = \left\{ (X,t) :\ A_{00}^k + \sum_{i=1}^m \sum_{j=1}^n A_{ij}^k X_{ij} + A_0^k t \succeq 0 \right\}.$$

Suppose in addition that there exist $X^s \in \mathbb{R}^{m\times n}$ and $s^s \in \mathbb{R}^p$ such that for $k=1,\ldots,p$,

$$A_{00}^k + \sum_{i=1}^m \sum_{j=1}^n A_{ij}^k X_{ij}^s + A_0^k s_k^s \succ 0.$$

If $\bar X \in \operatorname{dom}\partial f$ satisfies $0 \in \operatorname{ri}\partial f(\bar X)$, then f satisfies the KL property at $\bar X$ with exponent $\frac{1}{2}$.

Remark 4.2 Similar to Theorem 4.2, the "ri-condition" here is also imposed on f itself, while such a condition is imposed on the F in (3.7) in Theorem 4.1.

Proof Let F be defined as in (3.9) with the matrix variable $Z \in \mathcal{S}^{n+m}$ partitioned as in (3.10). Then $f(X) = \inf_{U,V} F(Z)$, thanks to (4.26). Let $r = \operatorname{rank}(\bar X)$ and

$$\bar X = [P_+\ P_0] \begin{bmatrix} \Sigma_+ & 0 \\ 0 & 0 \end{bmatrix} [Q_+\ Q_0]^T = P_+ \Sigma_+ Q_+^T,$$

be a singular value decomposition of $\bar X$, where $\Sigma_+ \in \mathbb{R}^{r\times r}$ is a diagonal matrix whose diagonal entries are the r positive singular values of $\bar X$, $[P_+\ P_0]$ is orthogonal with $P_+ \in \mathbb{R}^{m\times r}$ and $P_0 \in \mathbb{R}^{m\times(m-r)}$, $[Q_+\ Q_0]$ is orthogonal with $Q_+ \in \mathbb{R}^{n\times r}$ and $Q_0 \in \mathbb{R}^{n\times(n-r)}$. Define[15]

$$\bar Z := \begin{bmatrix} P_+\Sigma_+P_+^T & \bar X \\ \bar X^T & Q_+\Sigma_+Q_+^T \end{bmatrix}.$$

Then $\bar Z \succeq 0$. Now, using [45, Theorem 23.8], the definition of F and [45, Corollary 6.6.2], we have

$$\operatorname{ri}\partial F(\bar Z) = \left\{ \frac{1}{2}\begin{bmatrix} I_m & \Lambda \\ \Lambda^T & I_n \end{bmatrix} + Y : \Lambda \in \operatorname{ri}\partial\left(\sum_{k=1}^p f_k\right)(\bar X),\ Y \in \operatorname{ri} N_{\mathcal{S}_+^{m+n}}(\bar Z) \right\}. \tag{4.27}$$

Next, since $0 \in \operatorname{ri}\partial f(\bar X)$ and the nuclear norm is continuous, we see from [45, Theorem 23.8] and [45, Corollary 6.6.2] that

$$0 \in \operatorname{ri}\partial f(\bar X) = \operatorname{ri}\partial\left(\sum_{k=1}^p f_k\right)(\bar X) + \operatorname{ri}\partial\|\bar X\|_*. \tag{4.28}$$

[15] When $r=0$, we set $\bar Z = 0 \in \mathcal{S}^{m+n}$.

Moreover, recall from [55, Example 2] and [45, Corollary 7.6.1] that

$$\operatorname{ri}\partial\|\bar{X}\|_* = \left\{ [P_+\ P_0] \begin{bmatrix} I_r & 0 \\ 0 & W \end{bmatrix} [Q_+\ Q_0]^T :\ W \in \mathbb{R}^{(m-r)\times(n-r)}, \|W\|_2 < 1 \right\}, \tag{4.29}$$

where $\|W\|_2$ is the operator norm of W, that is, the largest singular value of W. Combining (4.28) and (4.29), we conclude that there exist $C \in \operatorname{ri}\partial\,(\sum_{k=1}^p f_k)(\bar{X})$ and W_0 with $\|W_0\|_2 < 1$ such that

$$0 = C + [P_+\ P_0] \begin{bmatrix} I_r & 0 \\ 0 & W_0 \end{bmatrix} [Q_+\ Q_0]^T = C + P_0 W_0 Q_0^T + P_+ Q_+^T. \tag{4.30}$$

On the other hand, using the definition of $\bar{Z}$ and a direct computation, we have

$$\bar{Z} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}}P_+ & P_0 & 0 & \frac{1}{\sqrt{2}}P_+ \\ \frac{1}{\sqrt{2}}Q_+ & 0 & Q_0 & -\frac{1}{\sqrt{2}}Q_+ \end{bmatrix}}_{\widehat{P}} \begin{bmatrix} 2\Sigma_+ & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}}P_+ & P_0 & 0 & \frac{1}{\sqrt{2}}P_+ \\ \frac{1}{\sqrt{2}}Q_+ & 0 & Q_0 & -\frac{1}{\sqrt{2}}Q_+ \end{bmatrix}^T. \tag{4.31}$$

Note that $\widehat{P}^T\widehat{P} = \widehat{P}\widehat{P}^T = I_{m+n}$, meaning that (4.31) is an eigenvalue decomposition of $\bar{Z}$. Thus, we can compute that

$$\begin{aligned} &\operatorname{ri} N_{\mathcal{S}_+^{m+n}}(\bar{Z}) = \operatorname{ri}\left[(-\mathcal{S}_+^{m+n}) \cap \{\bar{Z}\}^\perp\right] = \widehat{P} \begin{bmatrix} 0 & 0 \\ 0 & -\operatorname{int}\mathcal{S}_+^{m+n-r} \end{bmatrix} \widehat{P}^T \\ &\ni \begin{bmatrix} \frac{1}{\sqrt{2}}P_+ & P_0 & 0 & \frac{1}{\sqrt{2}}P_+ \\ \frac{1}{\sqrt{2}}Q_+ & 0 & Q_0 & -\frac{1}{\sqrt{2}}Q_+ \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2}I_{m-r} & \frac{1}{2}W_0 & 0 \\ 0 & \frac{1}{2}W_0^T & -\frac{1}{2}I_{n-r} & 0 \\ 0 & 0 & 0 & -I_r \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}}P_+^T & \frac{1}{\sqrt{2}}Q_+^T \\ P_0^T & 0 \\ 0 & Q_0^T \\ \frac{1}{\sqrt{2}}P_+^T & -\frac{1}{\sqrt{2}}Q_+^T \end{bmatrix} \\ &= \frac{1}{2}\begin{bmatrix} -I_m & -C \\ -C^T & -I_n \end{bmatrix}, \end{aligned}$$

where the inclusion holds because $\|W_0\|_2 < 1$, and the last equality follows from (4.30) and a direct computation. This together with (4.27) and the definition of C implies that $0 \in \operatorname{ri}\partial F(\bar{Z})$. Moreover, one can see that F is the sum of $p+1$ proper closed LMI-representable functions and the Slater's condition required in Theorem 4.2 holds. Thus, we conclude from Theorem 4.2 that F in (3.9) has KL property at $\bar{Z}$ with exponent $\frac{1}{2}$.

Finally, recall that for the F defined in (3.9), we have

$$\inf_{U,V} F(Z) = f(X) \text{ and } \operatorname*{Arg\,min}_{U,V} F\left(\begin{bmatrix} U & \bar{X} \\ \bar{X}^T & V \end{bmatrix}\right) = \left\{(P_+\Sigma_+P_+^T, Q_+\Sigma_+Q_+^T)\right\}.^{16}$$

These together with Corollary 3.3 and the fact that the KL exponent of F at $\bar{Z}$ is $\frac{1}{2}$ shows that f satisfies the KL property at $\bar{X}$ with exponent $\frac{1}{2}$. □

[16] When $r = 0$, this set is $\{(0,0)\}$ and $\bar{Z} = 0$.

Remark 4.3 In [60, Proposition 12], it was shown that if $\ell : \mathbb{R}^p \to \mathbb{R}$ is strongly convex on any compact convex set with locally Lipschitz gradient and $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a linear map, then the function

$$f(X) = \ell(\mathcal{A}X) + \|X\|_*$$

satisfies the KL property with exponent $\frac{1}{2}$ at any $\bar{X}$ that verifies $0 \in \operatorname{ri} \partial f(\bar{X})$. In particular, the loss function $X \mapsto \ell(\mathcal{A}X)$ is smooth. The more general case where the nuclear norm is replaced by a general spectral function was considered in [18, Theorem 3.12], and a sufficient condition involving the relative interior of *the subdifferential of the conjugate of the spectral function* was proposed in [18, Proposition 3.13], which, in general, is different from the regularity condition $0 \in \operatorname{ri} \partial f(\bar{X})$.

On the other hand, using our Theorem 4.3, we can deduce the KL exponent of functions in the form of (3.8) at points $\bar{X}$ satisfying the condition $0 \in \operatorname{ri} \partial f(\bar{X})$, but with a different set of conditions on the loss function. For instance, one can prove using Theorem 4.3 that the following functions f satisfy the KL property with exponent $\frac{1}{2}$ at a point $\bar{X}$ verifying $0 \in \operatorname{ri} \partial f(\bar{X})$:

(i) $f(X) = \frac{1}{2}\|\mathcal{A}X - b\|^2 + \mu \sum_{i,j} |X_{ij}| + \nu\|X\|_*$, where $\mu > 0$ and $\nu > 0$, $b \in \mathbb{R}^p$ and $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a linear map.
(ii) $f(X) = \|\mathcal{A}X - b\| + \mu \sum_{i,j} |X_{ij}| + \nu\|X\|_*$, where $\mu > 0$ and $\nu > 0$, $b \in \mathbb{R}^p$ and $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a linear map.

In view of [18, Theorem 3.12], it would be of interest to extend Theorem 4.3 to cover more general spectral functions. However, since our analysis in this subsection is based on LMI or SDP representability, it is not clear how this can be achieved at this moment. This would be a potential important future research direction.

Remark 4.4 (Discussion of the relative interior conditions) In Theorems 4.1, 4.2 and 4.3, the conclusions of KL exponent being 1/2 were derived under relative interior conditions. If these relative interior conditions were dropped, then the corresponding conclusions could fail, in general. For example, in [60, equation (53)], the authors provided an example of $\widetilde{f}(X) := f_1(X) + \|X\|_*$ for $X \in \mathbb{R}^{2\times 2}$, where f_1 is a convex quadratic function on $\mathbb{R}^{2\times 2}$, and showed that $0 \notin \operatorname{ri} \partial \widetilde{f}(\overline{X})$ for some $\overline{X} \in \mathbb{R}^{2\times 2}$ and the first-order error bound is not satisfied at $\overline{X}$. Recalling [14, Theorem 5] and [23, Corollary 3.6], this means that $\widetilde{f}$ cannot have a KL exponent of $\frac{1}{2}$ at $\overline{X}$.

We also would like to point out that, when the relative interior condition fails, one can follow the approach in Section 4.1 and the general error bound result for ill-posed semidefinite programs [24, 50] to derive a KL exponent that depends on the degree of singularity of a certain semidefinite system in the lifted representation. In general, this KL exponent will approach 1 quickly as the dimension grows, which can be of less interest. For simplicity, we do not discuss this in detail.

4.4 Convex models with C^2-cone reducible structure

In this section, we explore the KL exponent of functions that involve C^2-cone reducible structures. Our first theorem concerns the sum of the *support* function of a C^2-cone reducible closed convex set and a specially structured smooth convex function. In the theorem, we will also make use of the so-called bounded linear regularity condition [8, Definition 5.6]. Recall that $\{\mathfrak{D}_1, \mathfrak{D}_2\}$ is said to be boundedly linearly regular at $\bar{x} \in \mathfrak{D}_1 \cap \mathfrak{D}_2$ if for any bounded neighborhood $\mathfrak{U}$ of $\bar{x}$, there exists $c > 0$ such that

$$\operatorname{dist}(x, \mathfrak{D}_1 \cap \mathfrak{D}_2) \leq c[\operatorname{dist}(x, \mathfrak{D}_1) + \operatorname{dist}(x, \mathfrak{D}_2)] \text{ for all } x \in \mathfrak{U}.$$

It is known that if $\mathfrak{D}_1$ and $\mathfrak{D}_2$ are both polyhedral, then $\{\mathfrak{D}_1, \mathfrak{D}_2\}$ is boundedly linearly regular at any $\bar{x} \in \mathfrak{D}_1 \cap \mathfrak{D}_2$; moreover, if $\mathfrak{D}_1$ is polyhedral and $\mathfrak{D}_1 \cap \operatorname{ri} \mathfrak{D}_2 \neq \emptyset$, then $\{\mathfrak{D}_1, \mathfrak{D}_2\}$ is also boundedly linearly regular at any $\bar{x} \in \mathfrak{D}_1 \cap \mathfrak{D}_2$; see [9, Corollary 3].

Theorem 4.4 (Composite convex models with C^2-cone reducible structure) *Let $\ell : \mathbb{Y} \to \mathbb{R}$ be a function that is strongly convex on any compact convex set and has locally Lipschitz gradient, $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ be a linear map, and $v \in \mathbb{X}$. Consider the function*

$$h(x) := \ell(\mathcal{A}x) + \langle v, x \rangle + \sigma_{\mathfrak{D}}(x)$$

with $\mathfrak{D}$ being a nonempty C^2-cone reducible closed convex set. Suppose $0 \in \partial h(\bar{x})$. Then, one has

$$\bar{x} \in N_{\mathfrak{D}}(-\mathcal{A}^* \nabla \ell(\mathcal{A}\bar{x}) - v).$$

If we assume in addition that $\{\mathcal{A}^{-1}\{\mathcal{A}\bar{x}\}, N_{\mathfrak{D}}(-\mathcal{A}^* \nabla \ell(\mathcal{A}\bar{x}) - v)\}$ is boundedly linearly regular at $\bar{x}$, then h satisfies the KL property at $\bar{x}$ with exponent $\frac{1}{2}$.

Proof Since $0 \in \partial h(\bar{x})$, we see from [46, Exercise 8.8] that

$$\bar{w} := -\mathcal{A}^* \nabla \ell(\mathcal{A}\bar{x}) - v \in \partial \sigma_{\mathfrak{D}}(\bar{x}) = \partial \delta_{\mathfrak{D}}^*(\bar{x}) = (\partial \delta_{\mathfrak{D}})^{-1}(\bar{x}),$$

where the last equality follows from [46, Proposition 11.3]. This implies $\bar{x} \in \partial \delta_{\mathfrak{D}}(\bar{w}) = N_{\mathfrak{D}}(\bar{w})$.

We now assume in addition the bounded linear regularity condition and prove the alleged KL property. First, since $\mathfrak{D}$ is a C^2-cone reducible closed convex set, there exists $\tilde{\rho} > 0$ and a mapping $\Theta : \mathbb{X} \to \mathbb{V}$ which is twice continuously differentiable on $B(\bar{w}, \tilde{\rho})$ and a closed convex pointed cone $K \subseteq \mathbb{V}$ such that $\Theta(\bar{w}) = 0$, $D\Theta(\bar{w})$ is onto and $\mathfrak{D} \cap B(\bar{w}, \tilde{\rho}) = \{w : \Theta(w) \in K\} \cap B(\bar{w}, \tilde{\rho})$.

Fix any $\rho \in (0, \tilde{\rho})$ so that $D\Theta(w)$ is onto whenever $w \in B(\bar{w}, \rho)$. Then, we have from [46, Exercise 10.7] that

$$N_{\mathfrak{D}}(w) = D\Theta(w)^* N_K(\Theta(w)) \quad \text{for all } w \in B(\bar{w}, \rho). \tag{4.32}$$

Now, fix any $\delta > 0$. Take $w \in \mathfrak{D} \cap B(\bar{w}, \rho)$ and $x \in N_{\mathfrak{D}}(w) \cap B(\bar{x}, \delta)$. Then $x = D\Theta(w)^* u_x$ for some $u_x \in N_K(\Theta(w))$ according to (4.32). For such a u_x, one can observe that

$$D\Theta(\bar{w})^* u_x \in D\Theta(\bar{w})^* N_K(\Theta(w)) \subseteq D\Theta(\bar{w})^* K^\circ = D\Theta(\bar{w})^* N_K(\Theta(\bar{w})) = N_{\mathfrak{D}}(\bar{w}),$$

where K° is the polar of K, the set inclusion follows from the definition of normal cone and the fact that K is a closed convex cone, the first equality holds because $\Theta(\bar{w}) = 0$ and the last equality follows from (4.32). Thus, for any $w \in \mathfrak{D} \cap B(\bar{w}, \rho)$ and $x \in N_{\mathfrak{D}}(w) \cap B(\bar{x}, \delta)$, we have

$$\begin{aligned}\operatorname{dist}(x, N_{\mathfrak{D}}(\bar{w})) &\le \|x - D\Theta(\bar{w})^* u_x\| = \|D\Theta(w)^* u_x - D\Theta(\bar{w})^* u_x\| \\ &\le L\|u_x\|\|w - \bar{w}\|,\end{aligned} \tag{4.33}$$

where L is the Lipschitz continuity modulus of $D\Theta$ over the set $B(\bar{w}, \rho)$, which is finite because Θ is twice continuously differentiable.

Next, for each $z \in B(\bar{w}, \rho)$, define the linear map

$$\mathcal{W}(z) = \big(D\Theta(z) D\Theta(z)^*\big)^{-1} D\Theta(z).$$

Then $\mathcal{W}$ is continuously differentiable on $B(\bar{w}, \rho)$ because Θ is twice continuously differentiable on $B(\bar{w}, \rho)$ with surjective gradient map. Moreover, for any $w \in \mathfrak{D} \cap B(\bar{w}, \rho)$ and $x \in N_{\mathfrak{D}}(w) \cap B(\bar{x}, \delta)$, it follows from the definition of u_x that $[\mathcal{W}(w)](x) = u_x$. Let M be the Lipschitz continuity modulus of $w \mapsto \mathcal{W}(w)$ on $B(\bar{w}, \rho)$, which is finite because $\mathcal{W}$ is continuously differentiable on $B(\bar{w}, \rho)$. Then we have for any $w \in \mathfrak{D} \cap B(\bar{w}, \rho)$ and $x \in N_{\mathfrak{D}}(w) \cap B(\bar{x}, \delta)$ that

$$\begin{aligned}\|u_x - u_{\bar{x}}\| &= \|[\mathcal{W}(w)](x) - [\mathcal{W}(\bar{w})](\bar{x})\| \\ &\le \|[\mathcal{W}(w)](x) - [\mathcal{W}(\bar{w})](x)\| + \|[\mathcal{W}(\bar{w})](x) - [\mathcal{W}(\bar{w})](\bar{x})\| \\ &\le M\|x\|\,\|w - \bar{w}\| + \|\mathcal{W}(\bar{w})\|\|x - \bar{x}\| \\ &\le M\rho(\|\bar{x}\| + \|x - \bar{x}\|) + \|\mathcal{W}(\bar{w})\|\|x - \bar{x}\|,\end{aligned}$$

where the last inequality follows from triangle inequality and the fact that $w \in B(\bar{w}, \rho)$. In particular, $\|u_x\| \le \|u_{\bar{x}}\| + M\rho(\|\bar{x}\| + \delta) + \|\mathcal{W}(\bar{w})\|\delta =: \kappa$. This together with (4.33) implies that

$$N_{\mathfrak{D}}(w) \cap B(\bar{x}, \delta) \subseteq N_{\mathfrak{D}}(\bar{w}) + \kappa L\,\|w - \bar{w}\| B(0, 1) \quad \text{for all } w \in B(\bar{w}, \rho).$$

This means that the mapping $w \mapsto N_{\mathfrak{D}}(w)$ is calm at $\bar{w}$ with respect to $\bar{x}$; see [21, Page 182]. Thus, according to [21, Theorem 3H.3], the mapping $x \mapsto (N_{\mathfrak{D}})^{-1}(x)$ is metrically subregular at $\bar{x}$ with respect to $\bar{w}$; see [21, Page 183] for the definition. Noting also that $\partial\sigma_{\mathfrak{D}} = (N_{\mathfrak{D}})^{-1}$ according to [46, Example 11.4], we then deduce from [3, Theorem 3.3] that there exist $\delta' \in (0, \delta)$ and $c_0 > 0$ such that

$$\begin{aligned}&\sigma_{\mathfrak{D}}(x) - \sigma_{\mathfrak{D}}(\bar{x}) - \langle \bar{w}, x - \bar{x}\rangle \\ &\ge c_0 \operatorname{dist}(x, (\partial\sigma_{\mathfrak{D}})^{-1}(\bar{w}))^2 = c_0 \operatorname{dist}(x, N_{\mathfrak{D}}(\bar{w}))^2\end{aligned} \tag{4.34}$$

whenever $\|x-\bar{x}\| \leq \delta'$. We now follow a similar line of argument used in [60, Theorem 2] and [23, Theorem 4.2] to show the desired conclusion. Observe that

$$\begin{aligned}
&\operatorname{Arg\,min} h = \{z : 0 \in \partial h(z)\}\\
&= \{z : \mathcal{A}z = \mathcal{A}\bar{x} \text{ and } -\mathcal{A}^*\nabla\ell(\mathcal{A}z) - v \in (N_{\mathfrak{D}})^{-1}(z)\}\\
&= \{z : \mathcal{A}z = \mathcal{A}\bar{x} \text{ and } z \in N_{\mathfrak{D}}(-\mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x}) - v)\}.
\end{aligned}$$

Then it follows that for any bounded convex neighborhood $\mathfrak{U}$ of $\bar{x}$ with $\mathfrak{U} \subseteq B(\bar{x}, \delta')$, there exists $c_1 > 0$ such that for any $z \in \mathfrak{U}$,

$$\begin{aligned}
&\operatorname{dist}(z, \operatorname{Arg\,min} h) = \operatorname{dist}(z, \mathcal{A}^{-1}\{\mathcal{A}\bar{x}\} \cap N_{\mathfrak{D}}(\bar{w}))\\
&\overset{\text{(a)}}{\leq} \alpha[\operatorname{dist}(z, \mathcal{A}^{-1}\{\mathcal{A}\bar{x}\}) + \operatorname{dist}(z, N_{\mathfrak{D}}(\bar{w}))]\\
&\overset{\text{(b)}}{\leq} \alpha[c_1\|\mathcal{A}\bar{x} - \mathcal{A}z\| + \operatorname{dist}(z, N_{\mathfrak{D}}(\bar{w}))]\\
&\overset{\text{(c)}}{\leq} \alpha\left[c_1\|\mathcal{A}\bar{x} - \mathcal{A}z\| + c_0^{-\frac{1}{2}}\sqrt{\sigma_{\mathfrak{D}}(z) - \sigma_{\mathfrak{D}}(\bar{x}) - \langle \bar{w}, z - \bar{x}\rangle}\right];
\end{aligned} \tag{4.35}$$

here, (a) holds for some $\alpha > 0$ because of the bounded linear regularity assumption, (b) holds for some $c_1 > 0$ thanks to the Hoffman error bound, and (c) follows from (4.34). Now, as ℓ is strongly convex on compact convex sets, there exists $\beta > 0$ such that for all $z \in \mathfrak{U}$, we have

$$\beta\|\mathcal{A}\bar{x} - \mathcal{A}z\|^2 \leq \ell(\mathcal{A}z) - \ell(\mathcal{A}\bar{x}) - \langle \mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x}), z - \bar{x}\rangle.$$

Combining this with (4.35), we have for any $z \in \mathfrak{U}$ that

$$\begin{aligned}
\operatorname{dist}(z, \operatorname{Arg\,min} h) &\leq \alpha\left(c_1\|\mathcal{A}\bar{x} - \mathcal{A}z\| + c_0^{-\frac{1}{2}}\sqrt{\sigma_{\mathfrak{D}}(z) - \sigma_{\mathfrak{D}}(\bar{x}) - \langle \bar{w}, z - \bar{x}\rangle}\right)\\
&\leq \alpha c_1\beta^{-\frac{1}{2}}\sqrt{\ell(\mathcal{A}z) - \ell(\mathcal{A}\bar{x}) - \langle \mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x}), z - \bar{x}\rangle}\\
&\quad + \alpha c_0^{-\frac{1}{2}}\sqrt{\sigma_{\mathfrak{D}}(z) - \sigma_{\mathfrak{D}}(\bar{x}) - \langle \bar{w}, z - \bar{x}\rangle}.
\end{aligned}$$

Note that $\sqrt{a} + \sqrt{b} \leq \sqrt{2}\sqrt{a+b}$ for $a, b \geq 0$, and

$$h(z) - h(\bar{x}) = \ell(\mathcal{A}z) - \ell(\mathcal{A}\bar{x}) - \langle \mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x}), z - \bar{x}\rangle + \sigma_{\mathfrak{D}}(z) - \sigma_{\mathfrak{D}}(\bar{x}) - \langle \bar{w}, z - \bar{x}\rangle.$$

Thus, there exists $c > 0$ such that $\operatorname{dist}(z, \operatorname{Arg\,min} h) \leq c\sqrt{h(z) - h(\bar{x})}$ for all $z \in \mathfrak{U}$. Combining this with [14, Theorem 5], we conclude that h satisfies the KL property at $\bar{x}$ with exponent $\frac{1}{2}$. □

As a corollary of the preceding theorem, we consider the KL exponent of a class of gauge regularized optimization problems. Recall that a convex function $\gamma : \mathbb{X} \to \mathbb{R} \cup \{\infty\}$ is called a gauge if it is nonnegative, positively homogeneous, and vanishes at the origin. It is clear that any norm is a gauge. In the next corollary, we make explicit use of the gauge structure and replace the relative interior condition in Theorem 4.4 by one involving the so-called polar gauge. Recall from [28, Proposition 2.1(iii)] that for a gauge γ, its polar can be given by $\gamma^\circ(x) = \sup_z\{\langle x, z\rangle : \gamma(z) \leq 1\}$; moreover, polar of norms are their corresponding dual norms.

Corollary 4.1 *Let f be defined as in* (3.11). *Suppose that* $0 \in \partial f(\bar{x})$ *and* $\gamma(\bar{x}) > 0$. *Then* $\gamma^\circ(-\mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x})-v) = 1$. *Suppose in addition that* $-\mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x})-v \in \operatorname{dom}\partial\gamma^\circ$ *and the following relative interior condition holds:*

$$\mathcal{A}^{-1}\{\mathcal{A}\bar{x}\} \cap \left(\bigcup_{\lambda>0} \lambda\left(\operatorname{ri}\partial\gamma^\circ(-\mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x})-v)\right)\right) \neq \emptyset. \tag{4.36}$$

Then f satisfies the KL property at $\bar{x}$ with exponent $\frac{1}{2}$.

Proof Since $0 \in \partial f(\bar{x})$, we see from [46, Exercise 8.8] that

$$\bar{w} := -\mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x}) - v \in \partial\gamma(\bar{x}).$$

Since we have from [28, Proposition 2.1(iv)] that $\gamma^* = \delta_{\mathfrak{C}}$ with $\mathfrak{C} = \{x : \gamma^\circ(x) \le 1\}$, we conclude from (2.1) that $\gamma^\circ(\bar{w}) \le 1$ and $\gamma(\bar{x}) = \langle \bar{x}, \bar{w}\rangle$. Since $\gamma(\bar{x}) > 0$, we also have from $\gamma(\bar{x}) = \langle \bar{x}, \bar{w}\rangle$ and [28, Proposition 2.1(iii)] that

$$1 = \frac{\langle \bar{x}, \bar{w}\rangle}{\gamma(\bar{x})} \le \sup_z\{\langle \bar{w}, z\rangle :\ \gamma(z) \le 1\} = \gamma^\circ(\bar{w}).$$

Thus, it holds that $\gamma^\circ(\bar{w}) = 1$.

Next, suppose in addition that $\bar{w} \in \operatorname{dom}\partial\gamma^\circ$ and (4.36) holds. Let $F(x,t)$ be defined as in (3.12). Observe that

$$F(x,t) = \ell(\tilde{\mathcal{A}}(x,t)) + \langle (v,1),(x,t)\rangle + \sigma_{\mathfrak{D}^\circ}(x,t)$$

where $\tilde{\mathcal{A}}(x,t) := \mathcal{A}x$ and $\mathfrak{D}^\circ$ is the polar of $\mathfrak{D}$, which is given by $\mathfrak{D}^\circ = \{(x,t) : \gamma^\circ(x) + t \le 0\}$ according to the proof of [45, Theorem 15.4]. From our assumption, the set $\{(x,t) :\ \gamma^\circ(x) \le t\}$ is a C^2-cone reducible closed convex set, which implies that $\mathfrak{D}^\circ$ is also C^2-cone reducible. Now, observe from [45, Theorem 23.7] that for any $(u,s) \in \operatorname{dom}\partial\gamma^\circ \times \mathbb{R}$ satisfying $\gamma^\circ(u) + s = 0$, we have

$$N_{\mathfrak{D}^\circ}(u,s) = \operatorname{cl}\left(\bigcup_{\lambda\ge 0}\lambda\big(\partial\gamma^\circ(u),1\big)\right),$$

which together with [45, Theorem 6.3] and [45, Corollary 6.8.1] gives

$$\operatorname{ri} N_{\mathfrak{D}^\circ}(u,s) = \bigcup_{\lambda>0}\lambda\big(\operatorname{ri}\partial\gamma^\circ(u),1\big).$$

Applying this relation with $(u,s) = (\bar{w}, -\gamma^\circ(\bar{w})) = (\bar{w},-1)$ together with the relative interior condition (4.36) shows that

$$\big(\mathcal{A}^{-1}\{\mathcal{A}\bar{x}\} \times \mathbb{R}\big) \cap \operatorname{ri} N_{\mathfrak{D}^\circ}(\bar{w},-1) \neq \emptyset.$$

In view of this and [9, Corollary 3], we obtain that $\{\big(\mathcal{A}^{-1}\{\mathcal{A}\bar{x}\}\times\mathbb{R}\big), N_{\mathfrak{D}^\circ}(\bar{w},-1)\}$ is boundedly linearly regular. It follows from Theorem 4.4 that F satisfies the KL property at $(\bar{x},\gamma(\bar{x}))$ with exponent $\frac{1}{2}$. Since $f(x) = \inf_{t\in\mathbb{R}} F(x,t)$, we see from Corollary 3.3 that f satisfies the KL property at $\bar{x}$ with exponent $\frac{1}{2}$. $\square$

While checking C^2-cone reducibility directly using the definition can be difficult, a sufficient condition related to standard constraint qualifications was given in [47, Proposition 3.2].[17] Specifically, let $K \subseteq \mathbb{Y}$ be a C^2-cone reducible closed convex set and $G : \mathbb{X} \to \mathbb{Y}$ be a twice continuously differentiable function. If $G(\bar{x}) \in K$ and G is nondegenerate at $\bar{x}$ in the sense that

$$DG(\bar{x})\mathbb{X} + \big(T_K(G(\bar{x})) \cap \big[-T_K(G(\bar{x}))\big]\big) = \mathbb{Y}, \tag{4.37}$$

then $G^{-1}(K)$ is a C^2-cone reducible set. In particular, if $g_1, \ldots, g_m$ are C^2 functions with $\{\nabla g_i(\bar{x}) : i \in I(\bar{x})\}$ being linearly independent, where $I(\bar{x}) := \{i : \ g_i(\bar{x}) = 0\}$, then the set $\{x : g_i(x) \leq 0, i = 1, \ldots, m\}$ is C^2-cone reducible at $\bar{x}$.

We will now present a few concrete examples of functions to which Theorem 4.4 and Corollary 4.1 can be applied, taking advantage of the aforementioned sufficient condition (4.37) for checking C^2-cone reducibility.

Example 4.2 Let $\ell : \mathbb{Y} \to \mathbb{R}$ be a function that is strongly convex on any compact convex set and has locally Lipschitz gradient, $\mathcal{A} : \mathbb{X} \to \mathbb{Y}$ be a linear map, and $v \in \mathbb{X}$.

(i) **(Entropy-like regularization)** Let $\mathbb{X} = \mathbb{R}^n$ and $\mathbb{Y} = \mathbb{R}^m$. Denote

$$p(x) = \begin{cases} \sum_{i=1}^n x_i \log(x_i) - (\sum_{i=1}^n x_i) \log(\sum_{i=1}^n x_i) & \text{if } x \in \mathbb{R}^n_+, \\ \infty & \text{else,} \end{cases}$$

with the convention that $0 \log 0 = 0$. This function is proper closed convex and arises in the study of maximum entropy optimization [46, Example 11.12]. We claim that $f(x) = \ell(\mathcal{A}x) + \langle v, x\rangle + p(x)$ satisfies the KL property with exponent $\frac{1}{2}$ at any stationary point $\bar{x}$. To see this, recall from [46, Example 11.12] that

$$p(x) = \sigma_{\mathfrak{D}}(x), \quad \textit{where} \quad \mathfrak{D} = \{x \in \mathbb{R}^n : g(x) \leq 0\},$$

and $g(x) = \log(\sum_{i=1}^n e^{x_i})$. Then we have from Theorem 4.4 that $-\mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x}) - v \in \mathfrak{D}$. Moreover, for all $x \in \mathfrak{D}$, $\nabla g(x) = (\frac{e^{x_1}}{\sum_{i=1}^n e^{x_i}}, \ldots, \frac{e^{x_n}}{\sum_{i=1}^n e^{x_i}}) \neq 0$. Thus, in view of the discussion preceding this example, $\mathfrak{D}$ is C^2-cone reducible. Finally, notice that for any $x \in \mathfrak{D}$, the set

$$N_{\mathfrak{D}}(x) = \begin{cases} \bigcup_{\lambda \geq 0} \lambda\{\nabla g(x)\} & \text{if } g(x) = 0, \\ \{0\} & \text{if } g(x) < 0, \end{cases}$$

is polyhedral, and hence, $\{\mathcal{A}^{-1}\{\mathcal{A}\bar{x}\}, N_{\mathfrak{D}}(-\mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x}) - v)\}$ is boundedly linearly regular [8, Corollary 5.26]. So, Theorem 4.4 implies that f satisfies the KL property with exponent $\frac{1}{2}$ at any stationary point $\bar{x}$.

[17] The quoted result is for C^1-cone reducibility. However, it is apparent from the proof how to adapt the result for C^2-cone reducibility.

(ii) **(Positive semidefinite cone constraints)** Let $\mathbb{X} = \mathcal{S}^n$ and $\mathbb{Y} = \mathbb{R}^m$. Using the C^2-cone reducibility of $\mathcal{S}^n_+$, one can see that $f(X) = \ell(\mathcal{A}X) + \langle V, X\rangle + \delta_{\mathcal{S}^n_+}(X)$ satisfies the KL property with exponent $\frac{1}{2}$ at any stationary point $\bar{X}$ when $\mathcal{A}^{-1}\{\mathcal{A}\bar{X}\} \cap \mathrm{ri}\left(N_{-\mathcal{S}^n_+}(-\mathcal{A}^*\nabla\ell(\mathcal{A}\bar{X}) - V)\right) \neq \emptyset$. We note that this result has also been derived in [19] via a different approach.

(iii) **(Schatten p-norm regularization)** Let $\mathbb{X} = \mathcal{S}^n$ and $\mathbb{Y} = \mathbb{R}^m$. Let $p \in [1, 2] \cup \{\infty\}$ and consider the following optimization model with Schatten p-norm regularization:

$$f(X) = \ell(\mathcal{A}X) + \langle V, X\rangle + \tau\|X\|_p \quad \text{for all } X \in \mathcal{S}^n,$$

where $\|X\|_p = \left(\sum_{i=1}^n |\lambda_i(X)|^p\right)^{\frac{1}{p}}$ and $\lambda_n(X) \geq \lambda_{n-1}(X) \geq \cdots \geq \lambda_1(X)$ are eigenvalues of X. The dual norm of $\|\cdot\|_p$ is the Schatten q-norm with $\frac{1}{p} + \frac{1}{q} = 1$ where $q \in \{1\} \cup [2, \infty]$. Let $g(\lambda_1, \ldots, \lambda_n) = \left(\sum_{i=1}^n |\lambda_i|^q\right)^{\frac{1}{q}}$. It can be directly verified that g is convex, symmetric and C^2-cone reducible. So, $\|X\|_q = g(\lambda(X))$ is also C^2-cone reducible [18, Proposition 3.2]. Thus, from Corollary 4.1, f satisfies the KL property with exponent $\frac{1}{2}$ at any nonzero stationary point $\bar{X}$ under the relative interior condition (4.36) with $\gamma(X) = \|X\|_p$.

5 KL exponents for some nonconvex models

5.1 Difference-of-convex functions

In this section, we study a relationship between the KL exponents of the difference-of-convex (DC) function f in (3.13) and the auxiliary function F in (3.14). In [36, Theorem 4.1], it was shown that if f in (3.13) satisfies the KL property at $\bar{x} \in \mathrm{dom}\,\partial f$ with exponent $\frac{1}{2}$ and P_2 has globally Lipschitz gradient, then F in (3.14) satisfies the KL property at $(\bar{x}, \nabla P_2(\mathcal{A}\bar{x})) \in \mathrm{dom}\,\partial F$ with exponent $\frac{1}{2}$. Here we study the converse implication as a corollary to Theorem 3.1.

Theorem 5.1 (KL exponent of DC functions) *Suppose that f and F are defined in* (3.13) *and* (3.14) *respectively. If F is a KL function with exponent $\alpha \in [0, 1)$, then f is a KL function with exponent α.*

Proof Let $\bar{x} \in \mathrm{dom}\,\partial f$. We will show that f satisfies the KL property at $\bar{x}$ with exponent α.

Note that we have $\mathrm{dom}\,\partial f = \mathrm{dom}\,\partial P_1$ thanks to [46, Corollary 10.9] and the fact that continuous convex functions are locally Lipschitz continuous. Hence, we actually have $\bar{x} \in \mathrm{dom}\,\partial P_1$.

Now, using [46, Exercise 8.8] and [46, Proposition 10.5], we have for any $\bar{\xi} \in \partial P_2(\mathcal{A}\bar{x})$ that

$$\partial F(\bar{x}, \bar{\xi}) = \begin{bmatrix} \partial P_1(\bar{x}) - \mathcal{A}^*\bar{\xi} \\ \partial P_2^*(\bar{\xi}) - \mathcal{A}\bar{x} \end{bmatrix} \supseteq \begin{bmatrix} \partial P_1(\bar{x}) - \mathcal{A}^*\bar{\xi} \\ 0 \end{bmatrix}. \tag{5.1}$$

where the inclusion follows from the fact that $\partial P_2^* = \partial P_2^{-1}$ (see [46, Proposition 11.3]). Since $\bar{x} \in \operatorname{dom} \partial P_1$, we see further from (5.1) that $\{\bar{x}\} \times \partial P_2(\mathcal{A}\bar{x}) \subseteq \operatorname{dom} \partial F$. Then condition (i) of Theorem 3.1 holds because one can show using (2.1) that $\operatorname{Arg\,min}_y F(\bar{x}, y) = \partial P_2(\mathcal{A}\bar{x})$. On the other hand, the assumption on KL property of F shows that condition (ii) of Theorem 3.1 holds. Now, it remains to prove that F is level-bounded in y locally uniformly in x before we can apply Theorem 3.1 to establish the desired KL property.

To this end, we will show that for any $x^* \in \mathbb{X}$ and $\beta \in \mathbb{R}$, the following set is bounded:

$$\{(x, y) : \ \|x - x^*\| \le 1, \ F(x, y) \le \beta\}. \tag{5.2}$$

Suppose to the contrary that the above set is unbounded for some x^* and β. Then there exists a sequence

$$\{(x^k, y^k)\} \subseteq \{(x, y) : \ \|x - x^*\| \le 1, \ F(x, y) \le \beta\} \tag{5.3}$$

with $\|y^k\| \to \infty$. Passing to a subsequence if necessary, we may assume without loss of generality that $x^k \to \tilde{x}$ for some $\tilde{x} \in B(x^*, 1)$ and that $\lim_k \frac{y^k}{\|y^k\|}$ exists. Denote this latter limit by d. Then $\|d\| = 1$. Next, using the definition of $\{(x^k, y^k)\}$ in (5.3) and the definition of F, we have for all sufficiently large k that

$$\beta \ge F(x^k, y^k) = P_1(x^k) - \langle \mathcal{A}x^k, y^k\rangle + P_2^*(y^k) \ge f(x^k) \tag{5.4}$$

$$\Rightarrow \frac{\beta}{\|y^k\|} \ge \frac{P_1(x^k)}{\|y^k\|} - \left\langle \mathcal{A}x^k, \frac{y^k}{\|y^k\|} \right\rangle + \frac{P_2^*(y^k)}{\|y^k\|}, \tag{5.5}$$

where the second inequality in (5.4) follows from the definition of Fenchel conjugate. Then we see in particular from (5.4) and the closedness of f that $\tilde{x} \in \operatorname{dom} f = \operatorname{dom} P_1$. Using this, the closedness of P_1 and the definition of d, we have upon passing to limit inferior in (5.5) that

$$\begin{aligned} 0 &\ge -\langle \mathcal{A}\tilde{x}, d\rangle + \liminf_{k\to\infty} \frac{P_2^*(y^k)}{\|y^k\|} \overset{\text{(a)}}{\ge} -\langle \mathcal{A}\tilde{x}, d\rangle + (P_2^*)^\infty(d) \\ &\overset{\text{(b)}}{=} -\langle \mathcal{A}\tilde{x}, d\rangle + \sigma_{\operatorname{dom} P_2}(d) = -\langle \mathcal{A}\tilde{x}, d\rangle + \sup_{x \in \operatorname{dom} P_2} \{\langle x, d\rangle\}, \end{aligned}$$

where (a) follows from [7, Theorem 2.5.1] and (b) follows from [7, Theorem 2.5.4]. Since $\operatorname{dom} P_2 = \mathbb{Y}$, we deduce from the above inequality that $d = 0$, which contradicts the fact that $\|d\| = 1$. Thus, we have shown that (5.2) is bounded for any $x^* \in \mathbb{X}$ and any $\beta \in \mathbb{R}$, which implies that F is level-bounded in y locally uniformly in x. This completes the proof. □

5.2 Bregman envelope

In this section, we discuss the KL exponent of the Bregman envelope (3.15) of a proper closed function. We consider the following assumption on ϕ in (3.16), which is general enough for the corresponding (3.15) to include the celebrated Moreau envelope and the forward-backward envelope introduced in [49] as special cases. Further comments on this assumption will be given in Remark 5.1 below.

Assumption 5.1 The function ϕ in (3.16) is twice continuously differentiable and there exists $a_1 > 0$ such that for all $x \in \mathbb{X}$,

$$\nabla^2\phi(x) - a_1\mathcal{I} \succeq 0; \tag{5.6}$$

here $\mathcal{I}$ is the identity map, and for a linear map $\mathcal{A} : \mathbb{X} \to \mathbb{X}$, $\mathcal{A} \succeq 0$ means it is positive semidefinite, i.e., $\mathcal{A} = \mathcal{A}^*$ and $\langle h, \mathcal{A}h\rangle \geq 0$ for all $h \in \mathbb{X}$.

Given a proper closed function f and a function ϕ satisfying Assumption 5.1, we first analyze the KL property of the following auxiliary function:

$$F(x, y) := f(y) + \mathfrak{B}_\phi(y, x) \tag{5.7}$$

with $\mathfrak{B}_\phi$ defined in (3.16). For this function, applying [46, Proposition 8.8] and [46, Proposition 10.5], we have the following formula for ∂F at any $x \in \mathbb{X}$ and $y \in \operatorname{dom} f$,

$$\partial F(x, y) = \begin{bmatrix} -\nabla^2\phi(x)(y - x) \\ \partial f(y) + \nabla\phi(y) - \nabla\phi(x) \end{bmatrix}. \tag{5.8}$$

This formula will be used repeatedly in our discussion below.

Lemma 5.1 *Let $f : \mathbb{X} \to \mathbb{R} \cup \{\infty\}$ be a KL function with exponent $\alpha \in [\frac{1}{2}, 1)$. Let F be defined in* (5.7) *with ϕ satisfying Assumption 5.1. Then F is a KL function with exponent α.*

Proof Thanks to [34, Lemma 2.1], it suffices to show that F satisfies the KL property at any point (x, y) with $0 \in \partial F(x, y)$. Let $(\bar{x}, \bar{y})$ be such that $0 \in \partial F(\bar{x}, \bar{y})$. Then in view of (5.8), we see that $0 \in \partial F(\bar{x}, \bar{y})$ implies that $\nabla^2\phi(\bar{x})(\bar{y} - \bar{x}) = 0$. Combining this with (5.6) we deduce that $\bar{y} = \bar{x}$.

Next, since f is a KL function with exponent α, there exist $c, \eta, \epsilon > 0$ such that

$$\frac{1}{c}\operatorname{dist}^{\frac{1}{\alpha}}(0, \partial f(y)) \geq f(y) - f(\bar{x}) \tag{5.9}$$

whenever $y \in B(\bar{x}, \epsilon) \cap \operatorname{dom} \partial f$ and $f(y) < f(\bar{x}) + \eta$. Since ϕ is twice continuously differentiable, by shrinking ϵ further if necessary, we see that there exists $b_1 > a_1$ with a_1 being as in (5.6) such that for any $(x, y) \in B((\bar{x}, \bar{x}), \epsilon)$, there exists $x_0 \in B(\bar{x}, \epsilon)$ so that

$$\begin{aligned} &\|\nabla\phi(y) - \nabla\phi(x)\| \leq b_1\|y - x\| \\ \text{and } &\langle y - x, \nabla\phi(y) - \nabla\phi(x)\rangle = \langle y - x, [\nabla^2\phi(x_0)](y - x)\rangle. \end{aligned}$$

To the second relation in the above display, apply Cauchy-Schwartz inequality to the left hand side and apply (5.6) to the right hand side to obtain $\|y-x\|\|\nabla\phi(x)-\nabla\phi(y)\|\geq a_1\|y-x\|^2$. Combining this with the first relation in the above display, we obtain that

$$b_1\|y-x\|\geq\|\nabla\phi(y)-\nabla\phi(x)\|\geq a_1\|y-x\|. \tag{5.10}$$

Now, combining (5.8) with [34, Lemma 2.2], we deduce that there exists $C_0>0$ such that for $(x,y)\in B((\bar{x},\bar{x}),\epsilon)$ with $y\in\operatorname{dom}\partial f$,

$$\begin{aligned}
&\operatorname{dist}^{\frac{1}{\alpha}}(0,\partial F(x,y))\\
&\geq C_0\left(\|\nabla^2\phi(x)(y-x)\|^{\frac{1}{\alpha}}+\inf_{\xi\in\partial f(y)}\|\xi+\nabla\phi(y)-\nabla\phi(x)\|^{\frac{1}{\alpha}}\right)\\
&\overset{\text{(a)}}{\geq} C_0\left(a_1^{\frac{1}{\alpha}}\|y-x\|^{\frac{1}{\alpha}}+(a_1b_1^{-1})^{\frac{1}{\alpha}}\inf_{\xi\in\partial f(y)}\|\xi+\nabla\phi(y)-\nabla\phi(x)\|^{\frac{1}{\alpha}}\right)\\
&\overset{\text{(b)}}{\geq} C_0\left(a_1^{\frac{1}{\alpha}}\|y-x\|^{\frac{1}{\alpha}}+(a_1b_1^{-1})^{\frac{1}{\alpha}}\inf_{\xi\in\partial f(y)}\eta_1\|\xi\|^{\frac{1}{\alpha}}-(a_1b_1^{-1})^{\frac{1}{\alpha}}\eta_2\|\nabla\phi(y)-\nabla\phi(x)\|^{\frac{1}{\alpha}}\right)\\
&\overset{\text{(c)}}{\geq} C_0\left(a_1^{\frac{1}{\alpha}}\|y-x\|^{\frac{1}{\alpha}}+(a_1b_1^{-1})^{\frac{1}{\alpha}}\inf_{\xi\in\partial f(y)}\eta_1\|\xi\|^{\frac{1}{\alpha}}-a_1^{\frac{1}{\alpha}}\eta_2\|y-x\|^{\frac{1}{\alpha}}\right)\\
&\geq C_1\left(\inf_{\xi\in\partial f(y)}\|\xi\|^{\frac{1}{\alpha}}+\|y-x\|^{\frac{1}{\alpha}}\right),
\end{aligned} \tag{5.11}$$

where (a) follows from (5.6) and the fact that $\left(\frac{a_1}{b_1}\right)^{\frac{1}{\alpha}}<1$, (b) follows from [34, Lemma 3.1] for some $\eta_1>0$ and $\eta_2\in(0,1)$, (c) follows from the first inequality in (5.10), and the last inequality holds with $C_1:=C_0\min\{(1-\eta_2)a_1^{\frac{1}{\alpha}},\eta_1(a_1b_1^{-1})^{\frac{1}{\alpha}}\}>0$.

Next, since $\nabla\phi$ is Lipschitz continuous on $B(\bar{x},\epsilon/2)$ with Lipschitz constant b_1 in view of (5.10), by shrinking ϵ further, we may assume $2b_1\epsilon^2<1$ and that for any $(x,y)\in B((\bar{x},\bar{x}),\epsilon)$,

$$\begin{aligned}
0\leq\mathfrak{B}_\phi(y,x)&=\phi(y)-\phi(x)-\langle\nabla\phi(x),y-x\rangle\\
&\leq\frac{b_1}{2}\|y-x\|^2\leq\frac{b_1}{2}(2\epsilon)^2<1,
\end{aligned} \tag{5.12}$$

where the first inequality follows from the convexity of ϕ. Combining this with (5.11), we deduce further that for $(x,y)\in B((\bar{x},\bar{x}),\epsilon)$ with $y\in\operatorname{dom}\partial f$ and $F(x,y)<F(\bar{x},\bar{x})+\eta$,

$$\begin{aligned}
&\operatorname{dist}^{\frac{1}{\alpha}}(0,\partial F(x,y))\geq C_1\left(\inf_{\xi\in\partial f(y)}\|\xi\|^{\frac{1}{\alpha}}+\left(2b_1^{-1}\mathfrak{B}_\phi(y,x)\right)^{\frac{1}{2\alpha}}\right)\\
&\overset{\text{(a)}}{\geq} C_1\left(\inf_{\xi\in\partial f(y)}\|\xi\|^{\frac{1}{\alpha}}+(2b_1^{-1})^{\frac{1}{2\alpha}}\mathfrak{B}_\phi(y,x)\right)\\
&\overset{\text{(b)}}{=} C_1c\left(\inf_{\xi\in\partial f(y)}c^{-1}\|\xi\|^{\frac{1}{\alpha}}+(2b_1^{-1})^{\frac{1}{2\alpha}}c^{-1}\mathfrak{B}_\phi(y,x)\right)
\end{aligned}$$

$$\overset{\text{(c)}}{\geq} C_2 \left(\inf_{\xi \in \partial f(y)} c^{-1} \|\xi\|^{\frac{1}{\alpha}} + \mathfrak{B}_\phi(y,x) \right) \overset{\text{(d)}}{\geq} C_2 \left(f(y) - f(\bar{x}) + \mathfrak{B}_\phi(y,x) \right)$$
$$= C_2 \left(F(x,y) - F(\bar{x},\bar{x}) \right)$$

where (a) holds because $\frac{1}{2\alpha} \leq 1$ and $\mathfrak{B}_\phi(y,x) < 1$, thanks to (5.12), the constant c for (b) comes from (5.9), (c) holds with $C_2 := C_1 c \min\{1, (2b_1^{-1})^{\frac{1}{2\alpha}} c^{-1}\}$, (d) follows from (5.9) because $(x,y) \in B((\bar{x},\bar{x}),\epsilon)$, $y \in \operatorname{dom} \partial f$ and $f(y) \leq F(x,y) < F(\bar{x},\bar{x}) + \eta = f(\bar{x}) + \eta$, and the last equality holds because $f(\bar{x}) = F(\bar{x},\bar{x})$. This completes the proof. □

We are now ready to analyze the KL property of the Bregman envelope F_ϕ in (3.15).

Theorem 5.2 (KL exponent of Bregman envelope) *Let $f : \mathbb{X} \to \mathbb{R} \cup \{\infty\}$ be a proper closed function with* $\inf f > -\infty$*. Suppose that ϕ satisfies Assumption 5.1 and that f is a KL function with exponent $\alpha \in [\frac{1}{2}, 1)$. Then F_ϕ defined in* (3.15) *is a KL function with exponent α.*

Proof Let F be defined as in (5.7). We will use Theorem 3.1 to deduce the KL exponent of F_ϕ from that of F. To this end, we need to check all the conditions required by Theorem 3.1.

First, we claim that F is level-bounded in y locally uniformly in x. To prove this, fix any $x_0 \in \mathbb{X}$ and $t \in \mathbb{R}$. Define

$$U_{x_0} := \{(x,y) : \ \|x - x_0\| \leq 1, \ F(x,y) \leq t\}.$$

Thus, it suffices to show that U_{x_0} is bounded. To this end, note that ϕ is strongly convex with modulus a_1 according to Assumption 5.1. We have from this and the definition of Bregman distance that for any $(x,y) \in U_{x_0}$,

$$\frac{a_1}{2}\|x - y\|^2 \leq \mathfrak{B}_\phi(y,x).$$

Since $\inf f > -\infty$ by assumption, we deduce further that for any $(x,y) \in U_{x_0}$,

$$\inf f + \frac{a_1}{2}\|x - y\|^2 \leq \inf f + \mathfrak{B}_\phi(y,x) \leq f(y) + \mathfrak{B}_\phi(y,x) = F(x,y) \leq t.$$

Since $x \in B(x_0, 1)$, we deduce from the above inequality that U_{x_0} is bounded. Thus, we have shown that F is level-bounded in y locally uniformly in x.

Next, using [46, Exercise 8.8], we have for any $x \in \operatorname{dom} \partial F_\phi$ and any $\bar{y} \in \operatorname{Arg\,min}_y F(x,y)$ that

$$0 \in \partial f(\bar{y}) + \nabla \mathfrak{B}_\phi(\cdot, x)(\bar{y}),$$

which implies that $\partial f(\bar{y}) \neq \emptyset$. This together with (5.8) implies that $\partial F(x,\bar{y}) \neq \emptyset$ for any such x and $\bar{y}$. In particular, condition (i) in Theorem 3.1 is satisfied.

Finally, note that condition (ii) in Theorem 3.1 is also satisfied thanks to Lemma 5.1. Thus, we deduce from Theorem 3.1 that F_ϕ satisfies the KL property with exponent α at any $x \in \operatorname{dom} \partial F_\phi$. □

Remark 5.1 The Bregman envelope (3.15) with ϕ satisfying Assumption 5.1 covers several envelopes studied in the literature.

(i) When $\phi(\cdot) = \frac{1}{2\lambda}\|\cdot\|^2$ with some $\lambda > 0$, the function F_ϕ in (3.15) becomes

$$F_\phi(x) = \inf_y \left\{ f(y) + \frac{1}{2\lambda}\|x-y\|^2 \right\} =: e_\lambda f(x).$$

This function is known as the Moreau envelope of f. In [34, Theorem 3.4], it was proved that if f is a convex KL function with exponent $\alpha \in (0, \frac{2}{3})$ that is continuous on $\operatorname{dom}\partial f$, then $e_\lambda f$ is a KL function with exponent $\max\left\{\frac{1}{2}, \frac{\alpha}{2-2\alpha}\right\}$. Here, without the convexity and continuity assumptions, we can obtain a tighter estimate on the KL exponent of $e_\lambda f$ via Theorem 5.2: if f is a KL function with exponent $\alpha \in [\frac{1}{2}, 1)$ and $\inf f > -\infty$, then $e_\lambda f$ is a KL function with exponent α.

(ii) If the function f in (3.15) takes the form $h+g$, where g is a proper closed function, and h is twice continuously differentiable with Lipschitz gradient whose modulus is less than $\frac{1}{\gamma}$, then the function $\phi(x) := \frac{1}{2\gamma}\|x\|^2 - h(x)$ is convex and satisfies Assumption 5.1. The forward-backward envelope ψ_γ of the function $f = h + g$ was defined in [49] as follows (see also the discussion in [35, Section 2]):

$$\psi_\gamma(x) = \inf_y \{h(y) + g(y) + \mathfrak{B}_\phi(y, x)\}.$$

In [35, Theorem 3.2], it was shown that if the first-order error bound condition (or error bound condition in the sense of Luo-Tseng) holds for $h + g$, with h being in addition analytic and g being in addition convex, continuous on $\operatorname{dom}\partial g$, subanalytic and bounded below, then ψ_γ is a KL function with exponent $\frac{1}{2}$. Here, in view of Theorem 5.2, we can deduce the KL exponent of ψ_γ without the convexity and (sub)analyticity assumptions: if $f = h + g$ is a KL function with exponent $\alpha \in [\frac{1}{2}, 1)$ and $\inf f > -\infty$, g is a proper closed function, and h is twice continuously differentiable with Lipschitz gradient whose modulus is less than $\frac{1}{\gamma}$, then ψ_γ is a KL function with exponent α.

5.3 Least squares loss function with rank constraint

In this section, we compute an explicit KL exponent of the function f in (3.17), which can be rewritten as an inf-projection as in (3.18). Now, observe further that one can relax the orthogonality constraint and introduce a penalty function without changing the optimal value in (3.18), i.e.,

$$f(X) = \inf_U \left\{ \underbrace{\frac{1}{2}\|\mathcal{A}X - b\|^2 + \frac{1}{2}\|U^TU - I_{m-k}\|_F^2 + \delta_{\tilde{\mathfrak{D}}}(X,U)}_{\tilde{f}(X,U)} + \delta_{\tilde{\mathfrak{B}}}(X,U) \right\}, \tag{5.13}$$

where

$$\tilde{\mathfrak{D}} := \{(X,U) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times(m-k)} :\ U^T X = 0\},$$
$$\tilde{\mathfrak{B}} := \{(X,U) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times(m-k)} : 0.5 I_{m-k} \preceq U^T U \preceq 2I_{m-k}\},$$

where $A \preceq B$ means the matrix $B - A$ is positive semidefinite. In view of (5.13), as another application of Theorem 3.1, we will deduce the KL exponent of f via that of $\tilde{f} + \delta_{\tilde{\mathfrak{B}}}$.

We start with the following result, which is of independent interest.

Theorem 5.3 *Let $h : \mathbb{X} \to \mathbb{R}$ and $G : \mathbb{X} \to \mathbb{Y}$ be continuously differentiable. Assume that $G^{-1}\{0\} \neq \emptyset$ and define the functions g and g_1 by*

$$g(x) := h(x) + \delta_{G^{-1}\{0\}}(x), \qquad g_1(x,\lambda) := h(x) + \langle \lambda, G(x)\rangle.$$

Let $\bar{x} \in \operatorname{dom}\partial g$ and suppose that the linear map $\nabla G(\bar{x}) : \mathbb{Y} \to \mathbb{X}$ is injective. Then the following statements hold:

- (i) *There exists $\epsilon > 0$ so that for each $x \in B(\bar{x},\epsilon)$, the function $\lambda \mapsto \|\nabla h(x) + \nabla G(x)\lambda\|$ has a unique minimizer.*
- (ii) *If g_1 satisfies the KL property at $(\bar{x},\lambda(\bar{x}))$ with exponent α, then g satisfies the KL property at $\bar{x}$ with exponent α, where $\lambda(\bar{x})$ is the unique minimizer of $\lambda \mapsto \|\nabla h(\bar{x}) + \nabla G(\bar{x})\lambda\|$.*

Proof We first prove (i). Since $\nabla G(\bar{x})$ is an injective linear map and $x \mapsto \nabla G(x)$ is continuous, there exists an $\epsilon > 0$ so that $\nabla G(x)$ is an injective linear map whenever $x \in B(\bar{x},\epsilon)$. Then statement (i) follows immediately because the function $\lambda \mapsto \|\nabla h(x) + \nabla G(x)\lambda\|$ is minimized if and only if the quantity $\|\nabla h(x) + \nabla G(x)\lambda\|^2$ is minimized, and this latter function is a strongly convex function in λ whenever $x \in B(\bar{x},\epsilon)$, thanks to the fact that $\nabla G(x)$ is an injective linear map from $\mathbb{Y}$ to $\mathbb{X}$.

We now prove (ii). Let $x \in B(\bar{x},\epsilon)$ and $\lambda(x)$ denote the unique minimizer of $\lambda \mapsto \|\nabla h(x) + \nabla G(x)\lambda\|$. Then $\lambda(x)$ is also the unique minimizer of $\lambda \mapsto \|\nabla h(x) + \nabla G(x)\lambda\|^2$. Using the first-order optimality condition, we see that $\lambda(x)$ has to satisfy the relation $\nabla G(x)^* (\nabla h(x) + \nabla G(x)\lambda(x)) = 0$, which gives

$$\lambda(x) = -(\nabla G(x)^* \nabla G(x))^{-1}(\nabla G(x)^* \nabla h(x));$$

here the inverse exists because $\nabla G(x)$ is injective. Since h and G are continuously differentiable, we conclude that λ is a continuous function on $B(\bar{x},\epsilon)$.

Since g_1 satisfies the KL property at $(\bar{x},\lambda(\bar{x}))$ with exponent α, there exist $a,\nu,c > 0$ such that whenever $(x,\lambda) \in B\left((\bar{x},\lambda(\bar{x})),\nu\right)$ and $g_1(\bar{x},\lambda(\bar{x})) < g_1(x,\lambda) < g_1(\bar{x},\lambda(\bar{x})) + a$, it holds that

$$\|\nabla g_1(x,\lambda)\| \geq c\left(g_1(x,\lambda) - g_1(\bar{x},\lambda(\bar{x}))\right)^{\alpha}. \tag{5.14}$$

Next, using [46, Exercise 8.8], for any $x \in B(\bar{x},\epsilon) \cap \operatorname{dom}\partial g$, we have

$$\partial g(x) = \nabla h(x) + N_{G^{-1}\{0\}}(x) \subseteq \nabla h(x) + \{\nabla G(x)\lambda :\ \lambda \in \mathbb{Y}\},$$

where the inclusion follows from [46, Corollary 10.50] and the injectivity of $\nabla G(x)$. This implies that for any $x \in B(\bar{x}, \epsilon) \cap \mathrm{dom}\, \partial g$,

$$\mathrm{dist}(0, \partial g(x)) \geq \inf_{\lambda} \|\nabla h(x) + \nabla G(x)\lambda\| = \|\nabla h(x) + \nabla G(x)\lambda(x)\|, \tag{5.15}$$

where the equality follows from the definition of $\lambda(x)$ as the unique minimizer.

On the other hand, we have for any $x \in \mathrm{dom}\, \partial g$ and any λ that

$$\nabla g_1(x, \lambda) = \begin{bmatrix} \nabla h(x) + \nabla G(x)\lambda \\ G(x) \end{bmatrix} = \begin{bmatrix} \nabla h(x) + \nabla G(x)\lambda \\ 0 \end{bmatrix}, \tag{5.16}$$

where the second equality holds because $G(x) = 0$ whenever $x \in \mathrm{dom}\, \partial g$. Combining (5.16) with (5.15), we then obtain for any $x \in B(\bar{x}, \epsilon) \cap \mathrm{dom}\, \partial g$ that

$$\mathrm{dist}(0, \partial g(x)) \geq \|\nabla g_1(x, \lambda(x))\|. \tag{5.17}$$

Now, choose $0 < \epsilon' < \min\{\epsilon, \frac{\nu}{\sqrt{2}}\}$ small enough so that when $x \in B(\bar{x}, \epsilon') \cap \mathrm{dom}\, \partial g$, we have $\|\lambda(x) - \lambda(\bar{x})\| \leq \frac{\nu}{\sqrt{2}}$; such an ϵ' exists thanks to the continuity of $\lambda(\cdot)$. This implies that $(x, \lambda(x)) \in B\left((\bar{x}, \lambda(\bar{x})), \nu\right)$ whenever $x \in B(\bar{x}, \epsilon') \cap \mathrm{dom}\, \partial g$. Therefore, for $x \in B(\bar{x}, \epsilon') \cap \mathrm{dom}\, \partial g$ with $g(\bar{x}) < g(x) < g(\bar{x}) + a$, we have $(x, \lambda(x)) \in B\left((\bar{x}, \lambda(\bar{x})), \nu\right)$ and

$$g_1(\bar{x}, \lambda(\bar{x})) = g(\bar{x}) < g(x) = g_1(x, \lambda(x)) < g(\bar{x}) + a = g_1(\bar{x}, \lambda(\bar{x})) + a.$$

For these x, combining (5.14) with (5.17), we have

$$\mathrm{dist}(0, \partial g(x)) \geq c\left(g_1(x, \lambda(x)) - g_1(\bar{x}, \lambda(\bar{x}))\right)^{\alpha} = c\,\left(g(x) - g(\bar{x})\right)^{\alpha},$$

where the equality holds because $G(x) = 0$ whenever $x \in \mathrm{dom}\, \partial g$. This completes the proof. □

We now make use of Theorem 5.3 to deduce the KL exponent of $\tilde{f} + \delta_{\tilde{\mathfrak{B}}}$ in (5.13) at points $(\bar{X}, \bar{U}) \in \mathrm{dom}\, \partial(\tilde{f} + \delta_{\tilde{\mathfrak{B}}})$ with $\bar{U}^T\bar{U} = I_{m-k}$. For notational simplicity, we write

$$\tau := mn + m(m-k) + n(m-k) - 1. \tag{5.18}$$

Lemma 5.2 *The function $\tilde{f} + \delta_{\tilde{\mathfrak{B}}}$ given in* (5.13) *satisfies the KL property with exponent $1 - \frac{1}{4 \cdot 9^{\tau}}$ at points $(\bar{X}, \bar{U}) \in \mathrm{dom}\, \partial(\tilde{f} + \delta_{\tilde{\mathfrak{B}}})$ with $\bar{U}^T\bar{U} = I_{m-k}$, where τ is given in* (5.18).

Proof Define the function $G : \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times(m-k)} \to \mathbb{R}^{(m-k)\times n}$ by $G(X, U) := U^T X$, one can rewrite $\tilde{f}$ as

$$\tilde{f}(X, U) = \frac{1}{2}\|\mathcal{A}X - b\|^2 + \frac{1}{2}\|U^T U - I_{m-k}\|_F^2 + \delta_{G^{-1}\{0\}}(X, U).$$

Now, for $X \in \mathbb{R}^{m\times n}$, $U \in \mathbb{R}^{m\times(m-k)}$ and $\Lambda \in \mathbb{R}^{(m-k)\times n}$, define

$$\tilde{f}_1(X, U, \Lambda) := \frac{1}{2}\|\mathcal{A}X - b\|^2 + \frac{1}{2}\|U^T U - I_{m-k}\|_F^2 + \mathrm{tr}(\Lambda^T U^T X).$$

Note that $\tilde{f}_1$ is a polynomial of degree 4 on $\mathbb{R}^\tau$ where τ is given in (5.18). We deduce from [20, Theorem 4.2] that $\tilde{f}_1$ is a KL function with exponent $1-\frac{1}{4\cdot 9^\tau}$.

Next, since $(\bar{X},\bar{U}) \in \operatorname{dom}\partial(\tilde{f}+\delta_{\tilde{\mathfrak{B}}})$ with $\bar{U}^T\bar{U} = I_{m-k}$, we see that $(\bar{X},\bar{U})$ lies in the interior of $\tilde{\mathfrak{B}}$. Thus, we have $(\bar{X},\bar{U}) \in \operatorname{dom}\partial\tilde{f}$. We will now check the conditions in Theorem 5.3 for the functions $\tilde{f}_1$ and $\tilde{f}$ (in place of g_1 and g, respectively) at $(\bar{X},\bar{U})$. Notice first that the functions $(X,U) \mapsto \frac{1}{2}\|\mathcal{A}X - b\|^2 + \frac{1}{2}\|U^TU - I_{m-k}\|_F^2$ and G are continuously differentiable, and $G^{-1}\{0\}$ is clearly nonempty. We next claim that the linear map $\nabla G(\bar{X},\bar{U})$ is injective. To this end, let $Y \in \ker \nabla G(\bar{X},\bar{U})$. Then, using the definition of the derivative mapping of G, for any $(H,K) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times(m-k)}$, we have

$$\begin{aligned}0 &= \langle (H,K), [\nabla G(\bar{X},\bar{U})](Y)\rangle = \langle [DG(\bar{X},\bar{U})](H,K), Y\rangle \\ &= \langle \bar{U}^T H + K^T\bar{X}, Y\rangle = \langle H, \bar{U}Y\rangle + \langle \bar{X}Y^T, K\rangle.\end{aligned}$$

Since H and K are arbitrary, we deduce that

$$\bar{U}Y = 0 \ \text{ and } \ \bar{X}Y^T = 0.$$

These together with $\bar{U}^T\bar{U} = I_{m-k}$ imply that $Y = 0$. Thus, it holds that $\ker(\nabla G(\bar{X},\bar{U})) = \{0\}$, i.e., $\nabla G(\bar{X},\bar{U})$ is an injective linear map. Now, using Theorem 5.3, we conclude that $\tilde{f}$ satisfies the KL property at $(\bar{X},\bar{U})$ with exponent $1-\frac{1}{4\cdot 9^\tau}$.

Finally, since $(\bar{X},\bar{U}) \in \operatorname{int}\tilde{\mathfrak{B}}$, one can verify directly from the definition that, at $(\bar{X},\bar{U})$, the KL exponent of $\tilde{f}+\delta_{\tilde{\mathfrak{B}}}$ is the same as that of $\tilde{f}$. This completes the proof. □

Now we are ready to compute the KL exponent of f in (3.17). Interestingly, the derived KL exponent can be determined explicitly in terms of the number of rows/columns of the matrix involved and the upper bound constant in the rank constraint.

Theorem 5.4 *The function f given in* (3.17) *is a KL function with exponent $1-\frac{1}{4\cdot 9^\tau}$, where τ is given in* (5.18).

Proof Notice that $f(X) = \inf_U(\tilde{f}+\delta_{\tilde{\mathfrak{B}}})(X,U)$ and that for any $X \in \operatorname{dom}\partial f$,

$$\operatorname*{Arg\,min}_U(\tilde{f}+\delta_{\tilde{\mathfrak{B}}})(X,U) = \{U:\ U^TX = 0 \text{ and } U^TU = I_{m-k}\}, \tag{5.19}$$

where $\tilde{f}+\delta_{\tilde{\mathfrak{B}}}$ is given in (5.13). We will check the conditions in Theorem 3.1 and apply the theorem to deducing the KL exponent of f.

First, the function $\tilde{f}+\delta_{\tilde{\mathfrak{B}}}$ is clearly proper and closed. Next, for any fixed X, the U with $(X,U) \in \tilde{\mathfrak{D}} \cap \tilde{\mathfrak{B}}$ satisfies $0.5I_{m-k} \preceq U^TU \preceq 2I_{m-k}$. This shows that $\tilde{f}+\delta_{\tilde{\mathfrak{B}}}$ is bounded in U locally uniformly in X. Furthermore, for any $X \in \operatorname{dom}\partial f$ and any $U \in \operatorname{Arg\,min}_U(\tilde{f}+\delta_{\tilde{\mathfrak{B}}})(X,U)$, we have using (5.19) and [46, Exercise 8.8] that

$$\partial(\tilde{f}+\delta_{\tilde{\mathfrak{B}}})(X,U) = (\mathcal{A}^*(\mathcal{A}X-b),0) + N_{\tilde{\mathfrak{D}}\cap\tilde{\mathfrak{B}}}(X,U) \neq \emptyset.$$

These together with (5.19) and Lemma 5.2 implies that the conditions required by Theorem 3.1 are satisfied. Applying Theorem 3.1, we conclude that f is a KL function of exponent $1-\frac{1}{4\cdot 9^\tau}$. □

6 Concluding remarks

In this paper, we show that the KL exponent is preserved via inf-projection, under mild assumptions. The result is then used for studying KL exponents of various convex and nonconvex models, including some SDP-representable functions, convex functions involving C^2-cone reducible structures, Bregman envelopes, and more specifically, the sum of the least squares loss function and the indicator function of matrices of rank at most k.

Although several important calculus rules have been developed in this manuscript and the previous work [34], the KL exponents of some commonly used nonconvex models are still unknown, such as the least squares loss function with ℓ_{1-2} regularization [57]. Estimating the exponents for these models is an interesting future research question. Another future research direction will be to look at how KL exponent behaves under other important operations such as taking the maximum of finitely many or the supremum of infinitely many functions, as discussed in Remark 3.1. Finally, notice that many of our results in this paper for convex models require the strict complementarity condition $0 \in \operatorname{ri} \partial f(x)$. It will be interesting to identify suitable assumptions (other than polyhedral settings) under which the strict complementarity condition can be relaxed, as discussed in Remark 4.4.

References

1. M. Ahn, J. S. Pang and J. Xin, Difference-of-convex learning: directional stationarity, optimality, and sparsity, SIAM J. Optim. 27 (2017), 1637–1665.
2. C. M. Alaíz, Á. Barbero and J. R. Dorronsoro, Group fused lasso, in Artificial Neural Networks and Machine Learning–ICANN 2013. Lecture Notes in Computer Science, vol 8131. (V. Mladenov, P. Koprinkova-Hristova, G. Palm, A. E. P. Villa, B. Appollini, N. Kasabov, eds) Springer, Berlin, Heidelberg, 2013, pp. 66–73.
3. F. J. Aragón Artacho and M. H. Geoffroy, Characterization of metric regularity of subdifferentials, J. Convex Anal. 15 (2008), 365–380.
4. H. Attouch and J. Bolte, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, Math. Program. 116 (2009), 5–16.
5. H. Attouch, J. Bolte, P. Redont and A. Soubeyran, Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka-Lojasiewicz inequality, Math. Oper. Res. 35 (2010), 438–457.
6. H. Attouch, J. Bolte and B. F. Svaiter, Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward-backward splitting, and regularized Gauss-Seidel methods, Math. Program. 137 (2013), 91–129.
7. A. Auslender and M. Teboulle, Asymptotic Cones and Functions in Optimization and Variational Inequalities, Springer, 2003.
8. H. H. Bauschke and J. M. Borwein, On projection algorithms for solving convex feasibility problems, SIAM Review 38 (1996), 367–426.
9. H. H. Bauschke, J. M. Borwein and W. Li, Strong conical hull intersection property, bounded linear regularity, Jameson's property (G), and error bounds in convex optimization, Math. Program. 86 (1999), 135–160.
10. H. H. Bauschke, P. L. Combettes and D. Noll, Joint minimization with alternating Bregman proximity operators, Pac. J. Optim. 2 (2006), 401–424.
11. A. Ben-Tal and A. Nemirovski, Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications, MPS-SIAM Series on Optimization, 2001.

12. J. Bolte, A. Daniilidis and A. Lewis, The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, SIAM J. Optim. 17 (2007), 1205–1223.
13. J. Bolte, A. Daniilidis, A. Lewis and M. Shiota, Clarke subgradients of stratifiable functions, SIAM J. Optim. 18 (2007), 556–572.
14. J. Bolte, T. P. Nguyen, J. Peypouquet and B. W. Suter, From error bounds to the complexity of first-order descent methods for convex functions, Math. Program. 165 (2017), 471–507.
15. J. Bolte, S. Sabach and M. Teboulle, Proximal alternating linearized minimization for nonconvex and nonsmooth problems, Math. Program. 146 (2014), 459–494.
16. J. Borwein and A. Lewis, Convex Analysis and Nonlinear Optimization, 2nd edition, Springer, 2006.
17. S. Boyd, N. Parikh, E. Chu, B. Peleato and J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trend. in Mach. Learn. 3 (2010), 1–122.
18. Y. Cui, C. Ding and X. Zhao, Quadratic growth conditions for convex matrix optimization problems associated with spectral functions, SIAM J. Optim. 27 (2017), 2332–2355.
19. Y. Cui, D. F. Sun and K. C. Toh, On the asymptotic superlinear convergence of the augmented Lagrangian method for semidefinite programming with multiple solutions, Preprint 2016. Available at https://arxiv.org/abs/1610.00875.
20. D. D'Acunto and K. Kurdyka, Explicit bounds for the Łojasiewicz exponent in the gradient inequality for polynomials, Ann. Polon. Math. 87 (2005), 51–61.
21. A. L. Dontchev and R. T. Rockafellar, Implicit Functions and Solution Mappings, Springer, New York, 2009.
22. D. Drusvyatskiy, A. D. Ioffe and A. S. Lewis, Nonsmooth optimization using Taylor-like models: error bounds, convergence, and termination criteria, Math. Program. 185 (2021), 357-383.
23. D. Drusvyatskiy and A. Lewis, Error bounds, quadratic growth, and linear convergence of proximal methods, Math. Oper. Res. 43 (2018), 919–948.
24. D. Drusvyatskiy, G. Li and H. Wolkowicz, A note on alternating projections for ill-posed semidefinite feasibility problems, Math. Program. 162 (2017), 537–548.
25. J. Fan, Comments on "wavelets in statistics: a review" by A. Antoniadis, J. Ital. Stat. Soc. 6 (1997), 131–138.
26. F. Fachinei and J.-S. Pang, Finite-Dimensional Variational Inequalities and Complementarity Problems, Springer, New York, 2003.
27. P. Frankel, G. Garrigos and J. Peypouquet, Splitting methods with variable metric for Kurdyka-Łojasiewicz functions and general convergence rates, J. Optim. Theory Appl. 165 (2015), 874–900.
28. M. P. Friedlander, I. Macêdo and T. K. Pong, Gauge optimization and duality, SIAM J. Optim. 24 (2014), 1999–2022.
29. J. W. Helton and J. Nie, Semidefinite representation of convex sets, Math. Program. 122 (2010), 21–64.
30. R. Jiang and D. Li, Novel reformulations and efficient algorithms for the generalized trust region subproblem, SIAM J. Optim. 29 (2019), 1603–1633.
31. K. Kurdyka, On gradients of functions definable in o-minimal structures, Ann. Inst. Fourier 48 (1998) 769–783.
32. G. Li, B. S. Mordukhovich and T. S. Pham, New fractional error bounds for polynomial systems with applications to Hölderian stability in optimization and spectral theory of tensors, Math. Program. 153 (2015), 333–362.
33. G. Li and T. K. Pong, Douglas-Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems, Math. Program. 159 (2016), 371–401.
34. G. Li and T. K. Pong, Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods, Found. Comput. Math. 18 (2018), 1199–1232.
35. T. Liu and T. K. Pong, Further properties of the forward-backward envelope with applications to difference-of-convex programming, Comput. Optim. Appl. 67 (2017), 489–520.

36. T. Liu, T. K. Pong and A. Takeda, A refined convergence analysis of pDCA$_e$ with applications to simultaneous sparse recovery and outlier detection, Comput. Optim. Appl. 73 (2019), 69–100.
37. H. Liu, W. Wu and A. M.-C. So, Quadratic optimization with orthogonality constraints: explicit Łojasiewicz exponent and linear convergence of line-search methods, ICML, 2016, pp. 1158–1167.
38. S. Łojasiewicz, Une propriété topologique des sous-ensembles analytiques réels, in Les Équations aux Dérivées Partielles, Éditions du Centre National de la Recherche Scientifique, Paris, 1963, pp. 87–89.
39. B. F. Lourenço, M. Muramatsu and T. Tsuchiya, Facial reduction and partial polyhedrality, SIAM J. Optim. 28 (2018), 2304–2326.
40. Z. Q. Luo, J. S. Pang and D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, 1996.
41. Z. Q. Luo and P. Tseng, Error bounds and convergence analysis of feasible descent methods: a general approach, Ann. Oper. Res. 46 (1993), 157–178.
42. N. Parikh and S. P. Boyd, Proximal algorithms, Found. Trends Optimiz. 1 (2013), 123–231.
43. G. Pataki, The geometry of semidefinite programming, in Handbook of semidefinite programming (H. Wolkowicz, R. Saigal, and L. Vandenberghe, eds.), Kluwer Acad. Publ., Boston, MA, 2000, pp. 29–65.
44. B. Recht, M. Fazel and P. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization, SIAM Review 52 (2010), 471–501.
45. R. T. Rockafellar, Convex Analysis, Princeton University Press, 1970.
46. R. T. Rockafellar and R. J.-B. Wets, Variational Analysis, Springer, 1998.
47. A. Shapiro, Sensitivity analysis of generalized equations, J. Math. Sci. 115 (2003), 2554–2565.
48. A. Shapiro and K. Scheinberg, Duality and optimality conditions, in Handbook of semidefinite programming, Kluwer Acad. Publ., Boston, MA, 2000, pp. 67–110.
49. L. Stella, A. Themelis and P. Patrinos, Forward-backward quasi-Newton methods for nonsmooth optimization problems, Comput. Optim. Appl. 67 (2017), 443–487.
50. J. F. Sturm, Error bounds for linear matrix inequalities, SIAM J. Optim. 10 (2000), 1228–1248.
51. H. Tuy, Convex Analysis and Global Optimization, Springer, 2nd edition, 2016.
52. L. Tunçel and H. Wolkowicz, Strong duality and minimal representations for cone optimization, Comput. Optim. Appl. 53 (2012), 619–648.
53. P. Tseng and S. Yun, A coordinate gradient descent method for nonsmooth separable minimization, Math. Program. 117 (2009), 387–423.
54. M. Udell, C. Horn, R. Zadeh and S. Boyd, Generalized low rank models, Found. Trends in Mach. Learn. 9 (2016), 1–118.
55. A. Watson, Characterization of the subdifferential of some matrix norms, Linear Algebra Appl. 170 (1992), 33–45.
56. B. Wen, X. Chen and T. K. Pong, A proximal difference-of-convex algorithm with extrapolation, Comput. Optim. Appl. 69 (2018), 297–324.
57. P. Yin, Y. Lou, Q. He and J. Xin, Minimization of ℓ_{1-2} for compressed sensing, SIAM J. Sci. Comput. 37 (2015), A536–A563.
58. M. Yue, Z. Zhou and A. M.-C. So, A family of inexact SQA methods for non-smooth convex minimization with provable convergence guarantees based on the Luo-Tseng error bound property, Math. Program. 174 (2019), 327–358.
59. C.-H. Zhang, Nearly unbiased variable selection under minimax concave penalty, Ann. Stat. 38 (2010), 894–942.
60. Z. Zhou and A. M.-C. So, A unified approach to error bounds for structured convex optimization problems, Math. Program. 165 (2017), 689–728.
61. Z. Zhou, Q. Zhang and A. M.-C. So, $\ell_{1,p}$-norm regularization: error bounds and convergence rate analysis of first-order methods, ICML 2015, 2015, pp. 1501–1510.