Deducing Kurdyka-Łojasiewicz exponent of optimization models

Ting Kei Pong
Department of Applied Mathematics
The Hong Kong Polytechnic University
Hong Kong

RIKEN Center for AIP
April 2019
(Joint work with Guoyin Li and Peiran Yu)

Motivating applications

Structured optimization problems:

- (Overlapping) Group lasso:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2}\|Ax - b\|^2 + \sum_{i=1}^{m} \nu_i \|x_{J_i}\|,$$

  where $A \in \mathbb{R}^{p \times n}$, $b \in \mathbb{R}^p$, $\bigcup_{i=1}^m J_i = \{1, \ldots, n\}$, all $\nu_i \geq 0$.
- Least squares with rank constraint:

$$\begin{array}{ll} \min\limits_{X \in \mathbb{R}^{m \times n}} & \frac{1}{2}\|\mathcal{A}X - b\|^2 \\ \text{s.t.} & \operatorname{rank}(X) \leq r, \end{array}$$

  where $\mathcal{A} : \mathbb{R}^{m \times n} \to \mathbb{R}^p$ is linear, $b \in \mathbb{R}^p$, r is positive integer.

First-order method

Consider

$$f(x) := h(x) + P(x),$$

where: h is continuously differentiable with Lipschitz gradient whose continuity modulus is $L > 0$, P is proper closed.

First-order method

Consider

$$f(x) := h(x) + P(x),$$

where: h is continuously differentiable with Lipschitz gradient whose continuity modulus is $L > 0$, P is proper closed.

Many algorithms: proximal gradient, Douglas-Rachford splitting, etc.

First-order method

Consider

$$f(x) := h(x) + P(x),$$

where: h is continuously differentiable with Lipschitz gradient whose continuity modulus is $L > 0$, P is proper closed.

Many algorithms: proximal gradient, Douglas-Rachford splitting, etc.

Proximal gradient algorithm.
Initialize x^0, set $\gamma \in (0, \frac{1}{L})$. For $k = 0, \ldots,$

$$x^{k+1} \in \mathrm{prox}_{\gamma P}\left(x^k - \gamma \nabla h(x^k)\right),$$

where

$$\mathrm{prox}_{\gamma P}(y) = \operatorname*{Arg\,min}_{x \in \mathbb{R}^n} \left\{ \frac{1}{2}\|x - y\|^2 + \gamma P(x) \right\}.$$

KL property & exponent

Definition: (Attouch et al. '10, Attouch et al. '13)
Let f be proper closed and $\alpha \in [0, 1)$.

- f is said to have the Kurdyka-Łojasiewicz (KL) property with exponent α at $\bar{x} \in \mathrm{dom}\,\partial f$ if there exist c, ν, $\epsilon > 0$ so that

$$c[f(x) - f(\bar{x})]^{\alpha} \leq \mathrm{dist}(0, \partial f(x))$$

whenever $x \in \mathrm{dom}\,\partial f$, $\|x - \bar{x}\| \leq \epsilon$ and $f(\bar{x}) < f(x) < f(\bar{x}) + \nu$.

KL property & exponent

Definition: (Attouch et al. '10, Attouch et al. '13)
Let f be proper closed and $\alpha \in [0, 1)$.

- f is said to have the Kurdyka-Łojasiewicz (KL) property with exponent α at $\bar{x} \in \operatorname{dom} \partial f$ if there exist c, ν, $\epsilon > 0$ so that

$$c[f(x) - f(\bar{x})]^\alpha \leq \operatorname{dist}(0, \partial f(x))$$

  whenever $x \in \operatorname{dom} \partial f$, $\|x - \bar{x}\| \leq \epsilon$ and $f(\bar{x}) < f(x) < f(\bar{x}) + \nu$.
- If f has the KL property at every $\bar{x} \in \operatorname{dom} \partial f$ with the same α, then f is said to be a KL function with exponent α.

KL property & exponent

Definition: (Attouch et al. '10, Attouch et al. '13)
Let f be proper closed and $\alpha \in [0, 1)$.

- f is said to have the Kurdyka-Łojasiewicz (KL) property with exponent α at $\bar{x} \in \operatorname{dom} \partial f$ if there exist c, ν, $\epsilon > 0$ so that

$$c[f(x) - f(\bar{x})]^{\alpha} \leq \operatorname{dist}(0, \partial f(x))$$

  whenever $x \in \operatorname{dom} \partial f$, $\|x - \bar{x}\| \leq \epsilon$ and $f(\bar{x}) < f(x) < f(\bar{x}) + \nu$.
- If f has the KL property at every $\bar{x} \in \operatorname{dom} \partial f$ with the same α, then f is said to be a KL function with exponent α.

Examples.

- Proper closed semialgebraic functions are KL functions with exponent $\alpha \in [0, 1)$. (Bolte et al. '07)

Prototypical local convergence results

Fact 1.
For proximal gradient algorithm and some of its variants:
Let $\{x^k\}$ be a bounded sequence generated. If f is a KL function with exponent α, then:

- if $\alpha = 0$, then $\{x^k\}$ converges finitely;
- if $\alpha \in (0, \frac{1}{2}]$, then $\{x^k\}$ converges locally linearly;
- if $\alpha \in (\frac{1}{2}, 1)$, then $\{x^k\}$ converges locally sublinearly.

Prototypical local convergence results

Fact 1.
For proximal gradient algorithm and some of its variants:
Let $\{x^k\}$ be a bounded sequence generated. If f is a KL function with exponent α, then:

- if $\alpha = 0$, then $\{x^k\}$ converges finitely;
- if $\alpha \in (0, \frac{1}{2}]$, then $\{x^k\}$ converges locally linearly;
- if $\alpha \in (\frac{1}{2}, 1)$, then $\{x^k\}$ converges locally sublinearly.

Holds also for proximal alternating minimization algorithm (Attouch et al. '10), Douglas-Rachford splitting method (Li, P. '16), etc., if f is replaced by a suitable potential function.

Existing results I

For nonsmooth objectives:

- If f is the maximum of m polynomials of degree at most d, then the KL exponent is $1 - \frac{1}{\max\{1,(d+1)(3d)^{n+m-2}\}}$. (Li et al. '15)
- A special quadratic minimization problem with matrix variables and orthogonality constraint has KL exponent $\frac{1}{2}$. (Liu et al. '15)

Existing results I

For nonsmooth objectives:

- If f is the maximum of m polynomials of degree at most d, then the KL exponent is $1 - \frac{1}{\max\{1,(d+1)(3d)^{n+m-2}\}}$. (Li et al. '15)
- A special quadratic minimization problem with matrix variables and orthogonality constraint has KL exponent $\frac{1}{2}$. (Liu et al. '15)
- Relationship with Hölder growth condition: (Bolte et al. '17)
  Let f be proper closed convex with $\mathrm{Arg\,min}\, f \neq \emptyset$. Then f has KL exponent $\alpha \in [0,1)$ if and only if $\forall\ \bar{x} \in \mathrm{Arg\,min}\, f$, $\exists\ c, \epsilon > 0$ so that

  $$\mathrm{dist}(x, \mathrm{Arg\,min}\, f) \leq c(f(x) - f(\bar{x}))^{1-\alpha}$$

  whenever $\|x - \bar{x}\| \leq \epsilon$ and $f(\bar{x}) < f(x) < f(\bar{x}) + \epsilon$.

Existing results I

For nonsmooth objectives:

- If f is the maximum of m polynomials of degree at most d, then the KL exponent is $1 - \frac{1}{\max\{1,(d+1)(3d)^{n+m-2}\}}$. (Li et al. '15)
- A special quadratic minimization problem with matrix variables and orthogonality constraint has KL exponent $\frac{1}{2}$. (Liu et al. '15)
- Relationship with Hölder growth condition: (Bolte et al. '17)
  Let f be proper closed convex with $\operatorname{Arg\,min} f \neq \emptyset$. Then f has KL exponent $\alpha \in [0, 1)$ if and only if $\forall\, \bar{x} \in \operatorname{Arg\,min} f$, $\exists\, c, \epsilon > 0$ so that

  $$\operatorname{dist}(x, \operatorname{Arg\,min} f) \leq c(f(x) - f(\bar{x}))^{1-\alpha}$$

  whenever $\|x - \bar{x}\| \leq \epsilon$ and $f(\bar{x}) < f(x) < f(\bar{x}) + \epsilon$.
- A convex piecewise polynomial function of degree at most d is a KL function with exponent $1 - \frac{1}{(d-1)^n+1}$. (Li '13)

Existing results II

For nonsmooth objectives:

- Relationship with Luo-Tseng error bound: (Li, P. '18)
  Let $f = h + P$, where h has locally Lipschitz gradient and P is proper closed convex. Suppose $\mathcal{X} := \{x : \ 0 \in \partial f(x)\} \neq \emptyset$, and
  1. $\forall\, \zeta \geq \inf f$, $\exists\, c, \epsilon > 0$ so that
  $$\mathrm{dist}(x, \mathcal{X}) \leq c \|\mathrm{prox}_P(x - \nabla h(x)) - x\|$$
  whenever $\|\mathrm{prox}_P(x - \nabla h(x)) - x\| < \epsilon$ and $f(x) \leq \zeta$.
  2. $\exists\, \delta > 0$ so that if $x, y \in \mathcal{X}$ and $\|x - y\| \leq \delta$, then $f(x) = f(y)$.

  Then f is a KL function with exponent $\frac{1}{2}$.

Existing results III

For nonsmooth objectives:
Consequently: If $f = h + P$ and $\mathcal{X} \neq \emptyset$, then f satisfies the KL property with exponent $\frac{1}{2}$ at $\bar{x} \in \mathcal{X}$ in each of the following cases:

- h is a quadratic (not necessarily convex) and P is proper polyhedral. (Luo, Tseng '92, Tseng, Yun '09)
- $h(x) = \ell(Ax)$, where $\ell \in C^2$ is strongly convex on any compact convex set, and
    1. P is proper polyhedral; (Luo, Tseng '92, Tseng, Yun '09)
    2. $P(x) = \sum_{i=1}^{m} w_i \|x_{J_i}\|_p$, $w_i \geq 0$, $\{J_1, \ldots, J_m\}$ form a partition of $\{1, \ldots, n\}$, $p \in [1, 2] \cup \{\infty\}$; (Tseng '10, Zhou et al. '15)

Existing results III

For nonsmooth objectives:

Consequently: If $f = h + P$ and $\mathcal{X} \neq \emptyset$, then f satisfies the KL property with exponent $\frac{1}{2}$ at $\bar{x} \in \mathcal{X}$ in each of the following cases:

- h is a quadratic (not necessarily convex) and P is proper polyhedral. (Luo, Tseng '92, Tseng, Yun '09)
- $h(x) = \ell(Ax)$, where $\ell \in C^2$ is strongly convex on any compact convex set, and
  1. P is proper polyhedral; (Luo, Tseng '92, Tseng, Yun '09)
  2. $P(x) = \sum_{i=1}^{m} w_i \|x_{J_i}\|_p$, $w_i \geq 0$, $\{J_1, \ldots, J_m\}$ form a partition of $\{1, \ldots, n\}$, $p \in [1, 2] \cup \{\infty\}$; (Tseng '10, Zhou et al. '15)
  3. P is the nuclear norm if in addition $0 \in \mathrm{ri}\,\partial f(\bar{x})$; (Zhou, So '17)
  4. $P(x) = g(\sigma(x))$, g is polyhedral symmetric, under some ri conditions. (Cui et al. '17)

Aim: Estimate KL exponent

Aim: Explicitly estimate the KL exponent of optimization models.

Strategy:

- Develop calculus rules on KL exponents:

Deduce exponent of functions from ones with known exponents

Calculus of KL exponent I

Theorem 1. (Li, P. '18)
Let $h(x) = \ell(Ax)$ for some continuous strictly convex function ℓ, g be a continuous convex function, D be a closed convex set, $\alpha \in (0, 1)$. Suppose also

(i) there exists $x_0 \in D$ with $g(x_0) < 0$;

(ii) $\inf_{x \in D} h(x) < \inf_{x \in D}\{h(x) : \; g(x) \leq 0\}$;

(iii) for any $\lambda > 0$, $h + \lambda g + \delta_D$ is KL with exponent α.

Then $h + \delta_{g(\cdot) \leq 0} + \delta_D$ is KL with exponent α.

Application I

Consider functions of the form

$$f(x) = \ell(Ax) + \delta_\Omega(x),$$

where $\ell \in C^2$ is strongly convex on any compact convex set, and

$$\Omega := \left\{ x : \sum_{i=1}^{m} w_i \|x_i\|_p \leq \sigma \right\},$$

with $x_i \in \mathbb{R}^{n_i}$, $\sum_{i=1}^{m} n_i = n$, $w_i > 0$, $\sigma > 0$ and $p \in [1, 2]$.

Application I

Consider functions of the form

$$f(x) = \ell(Ax) + \delta_\Omega(x),$$

where $\ell \in C^2$ is strongly convex on any compact convex set, and

$$\Omega := \left\{ x : \sum_{i=1}^{m} w_i \|x_i\|_p \le \sigma \right\},$$

with $x_i \in \mathbb{R}^{n_i}$, $\sum_{i=1}^m n_i = n$, $w_i > 0$, $\sigma > 0$ and $p \in [1, 2]$.

Corollary 1. (Li, P. '18)
Suppose that $\inf f(x) > \inf \ell(Ax)$. Then f is KL with exponent $\frac{1}{2}$.

Application I

Consider functions of the form

$$f(x) = \ell(Ax) + \delta_\Omega(x),$$

where $\ell \in C^2$ is strongly convex on any compact convex set, and

$$\Omega := \left\{ x : \sum_{i=1}^{m} w_i \|x_i\|_p \le \sigma \right\},$$

with $x_i \in \mathbb{R}^{n_i}$, $\sum_{i=1}^m n_i = n$, $w_i > 0$, $\sigma > 0$ and $p \in [1, 2]$.

Corollary 1. (Li, P. '18)
Suppose that $\inf f(x) > \inf \ell(Ax)$. Then f is KL with exponent $\frac{1}{2}$.

Proof: When $\mathcal{X} \neq \emptyset$, Luo-Tseng error bound holds for the regularized version. (Zhou et al. '15)

Calculus of KL exponent II

Theorem 2. (Li, P. '18, Yu, Li, P. '19)
Suppose that g_i are KL functions with exponents α_i, $1 \leq i \leq m$, and that $\mathrm{dom}\,\partial g_i = \mathrm{dom}\, g_i$ for all i. Then $g := \min_{1\leq i\leq m} g_i$ is a KL function with exponent $\max\{\alpha_i : \ 1 \leq i \leq m\}$.

Calculus of KL exponent II

Theorem 2. (Li, P. '18, Yu, Li, P. '19)
Suppose that g_i are KL functions with exponents α_i, $1 \le i \le m$, and that $\mathrm{dom}\,\partial g_i = \mathrm{dom}\, g_i$ for all i. Then $g := \min_{1 \le i \le m} g_i$ is a KL function with exponent $\max\{\alpha_i : \ 1 \le i \le m\}$.

Key fact used in the proof: For any $x \in \mathrm{dom}\,\partial g$,

$$\partial g(x) \subseteq \bigcup_{i \in I(x)} \partial g_i(x),$$

where $I(x) := \{i : \ g(x) = g_i(x)\}$. (Mordukovich, Shao '95)

Application II

Corollary 2. (Li, P. '18, Yu, Li, P. '19)
Consider functions of the form

$$f(x) = \min_{1 \le i \le m} \left\{ x^T M_i x + b_i^T x + c_i + P_i(x) \right\},$$

where M_i are symmetric matrices, P_i are proper polyhedral functions. Then f is a KL function with exponent $\frac{1}{2}$.

Application II

Corollary 2. (Li, P. '18, Yu, Li, P. '19)
Consider functions of the form

$$f(x) = \min_{1\le i\le m}\left\{x^T M_i x + b_i^T x + c_i + P_i(x)\right\},$$

where M_i are symmetric matrices, P_i are proper polyhedral functions.
Then f is a KL function with exponent $\frac{1}{2}$.

Example: Least-squares with SCAD regularization: (Fan '97)

$$f(x) = \frac{1}{2}\|Ax - b\|^2 + \sum_{i=1}^{n} r_{\lambda,\theta}(x_i),$$

with $\lambda > 0$, $\theta > 2$ and

$$r_{\lambda,\theta}(t) = \begin{cases} \lambda|t| & \text{if } |t| \le \lambda, \\ \frac{-t^2+2\theta\lambda|t|-\lambda^2}{2(\theta-1)} & \text{if } \lambda < |t| \le \theta\lambda, \\ \frac{(\theta+1)\lambda^2}{2} & \text{if } |t| > \theta\lambda. \end{cases}$$

Calculus of KL exponent III

Theorem 3. (Yu, Li, P. '19)
Let $F : \mathbb{X} \times \mathbb{Y} \to \bar{\mathbb{R}}$ be proper closed 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 $\bar{x} \in \operatorname{dom} \partial f$, $\alpha \in [0, 1)$ and the following conditions hold:

(i) F is level-bounded in y locally uniformly in x.

(ii) It holds that $\partial F(\bar{x}, \bar{y}) \neq \emptyset$ for all $\bar{y} \in Y(\bar{x})$.

(iii) F satisfies the KL property with exponent α in $\{\bar{x}\} \times Y(\bar{x})$.

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

Remark: F is level-bounded in y locally uniformly in x if for any x and $\beta \in \mathbb{R}$, there exists $\rho > 0$ so that

$$\{(u, y) : \ \|u - x\| \leq \rho, F(u, y) \leq \beta\}$$

is bounded.

Application III

Corollary 3. (Yu, Li, P. '19)
Let $f = \sum_{i=1}^m f_i$, each $f_i : \mathbb{R}^n \to \bar{\mathbb{R}}$ be proper closed.
Suppose that each f_i is LMI-representable, i.e., there exist $d_i > 0$ and matrices $\{A_{00}^i, A_0^i, A_1^i, \ldots, A_n^i\} \subset \mathcal{S}^{d_i}$ such that

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

Suppose also that $\exists\ x^s \in \mathbb{R}^n$ 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 $0 \in \mathrm{ri}\,\partial f(\bar{x})$, then f satisfies the KL property at $\bar{x}$ with exponent $\frac{1}{2}$.

Application III cont.

Each of the following functions satisfies the KL property with exponent $\frac{1}{2}$ at an $\bar{x}$ satisfying $0 \in \mathrm{ri}\,\partial f(\bar{x})$:

(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 $b \in \mathbb{R}^p$, $A \in \mathbb{R}^{p\times n}$, $\bigcup_{i=1}^{s} J_i = \{1, \ldots, n\}$, all $w_i \geq 0$, $i = 1, \ldots, s$.

(ii) Group fused Lasso: (Alaíz et al. '13)

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

where $b \in \mathbb{R}^p$, $A \in \mathbb{R}^{p\times sr}$, $\bigcup_{i=1}^{s} J_i = \{1, \ldots, n\}$, $J_i \cap J_{i'} = \emptyset$ for $i \neq i'$, all $w_i, \nu_i \geq 0$ and $|J_i| = r$.

Application IV

Corollary 4. (Yu, Li, P. '19)
Let $f(X) = \sum_{k=1}^m f_k(X) + \|X\|_*$, each $f_k : \mathbb{R}^{m\times n} \to \bar{\mathbb{R}}$ be proper closed. Suppose that each f_k is LMI-representable, i.e., there exist $d_k > 0$ and matrices $\{A_{00}^k, A_0^k, A_{11}^k, \ldots, A_{mn}^k\} \subset \mathcal{S}^{d_k}$ such that

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

Suppose also that $\exists\, X^s \in \mathbb{R}^{m\times n}$ and $s^s \in \mathbb{R}^m$ such that for $k = 1, \ldots, m$,

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

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

Digression: C^2-cone reducible structures

Definition: (Shapiro '03)
Let $\mathfrak{D} \subseteq \mathbb{X}$ be a nonempty closed set. We say that it 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 $\Theta : \mathbb{X} \to \mathbb{Y}$ such that

(i) Θ is twice continuously differentiable in $B(\bar{w}, \rho)$;

(ii) $\Theta(\bar{w}) = 0$ and $D\Theta(\bar{w}) : \mathbb{X} \to \mathbb{Y}$ is onto;

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

We say that $\mathfrak{D}$ is C^2-cone reducible if it is C^2-cone reducible at every $\bar{w} \in \mathfrak{D}$.

Digression: C^2-cone reducible structures

Definition: (Shapiro '03)
Let $\mathfrak{D} \subseteq \mathbb{X}$ be a nonempty closed set. We say that it 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 $\Theta : \mathbb{X} \to \mathbb{Y}$ such that

(i) Θ is twice continuously differentiable in $B(\bar{w}, \rho)$;

(ii) $\Theta(\bar{w}) = 0$ and $D\Theta(\bar{w}) : \mathbb{X} \to \mathbb{Y}$ is onto;

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

We say that $\mathfrak{D}$ is C^2-cone reducible if it is C^2-cone reducible at every $\bar{w} \in \mathfrak{D}$.

Examples: (Shapiro '03)

1. Polyhedral sets, second-order cone, positive semidefinite cone.

Digression: C^2-cone reducible structures

Definition: (Shapiro '03)
Let $\mathfrak{D} \subseteq \mathbb{X}$ be a nonempty closed set. We say that it 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 $\Theta : \mathbb{X} \to \mathbb{Y}$ such that

(i) Θ is twice continuously differentiable in $B(\bar{w}, \rho)$;

(ii) $\Theta(\bar{w}) = 0$ and $D\Theta(\bar{w}) : \mathbb{X} \to \mathbb{Y}$ is onto;

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

We say that $\mathfrak{D}$ is C^2-cone reducible if it is C^2-cone reducible at every $\bar{w} \in \mathfrak{D}$.

Examples: (Shapiro '03)

1. Polyhedral sets, second-order cone, positive semidefinite cone.
2. $\mathfrak{D} = \{x : \ g_i(x) \leq 0, i = 1, \ldots, m\}$, $g_i \in C^2$, LICQ holds at $\bar{x} \in \mathfrak{D}$ $\Rightarrow$ C^2-cone reducible at $\bar{x}$.

Digression: C^2-cone reducible structures

Theorem 4. (Yu, Li, P. '19)
Let $\ell : \mathbb{Y} \to \mathbb{R}$ be strongly convex on any compact convex set and have 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 C^2-cone reducible closed convex set. Suppose that $0 \in \partial h(\bar{x})$ and

$$\mathcal{A}^{-1}\{\mathcal{A}\bar{x}\} \cap \mathrm{ri} N_{\mathfrak{D}}(-\mathcal{A}^*\nabla\ell(\mathcal{A}\bar{x}) - v) \neq \emptyset,$$

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

Application V

Let $\ell : \mathbb{R}^m \to \mathbb{R}$ be strongly convex on any compact convex set and have locally Lipschitz gradient, $\mathcal{A} : \mathcal{S}^n \to \mathbb{R}^m$ be a linear map.

Each of the following functions satisfies the KL property with exponent $\frac{1}{2}$ at an $\bar{X}$ satisfying the **ri condition**:

(i) PSD cone constraint:

$$f(X) = \ell(\mathcal{A}X) + \langle V, X\rangle + \delta_{\mathcal{S}^n_+}(X).$$

(ii) Schatten p-norm regularization:

$$f(X) = \ell(\mathcal{A}X) + \langle V, X\rangle + \mu\|X\|_p,$$

where $\|X\|_p$ is the Schatten p-norm with $p \in [1, 2] \cup \{\infty\}$ and $\bar{X} \neq 0$.

Application VI

Corollary 5. (Yu, Li, P. '19)
Let f be proper closed with $\inf f > -\infty$ and $\phi \in C^2$ is strongly convex. If f is KL with exponent $\alpha \in [\frac{1}{2}, 1)$, then so is the envelope function

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

where $\mathfrak{B}_\phi$ is the Bregman distance:

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

Application VI

Corollary 5. (Yu, Li, P. '19)
Let f be proper closed with $\inf f > -\infty$ and $\phi \in C^2$ is strongly convex. If f is KL with exponent $\alpha \in [\frac{1}{2}, 1)$, then so is the envelope function

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

where $\mathfrak{B}_\phi$ is the Bregman distance:

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

Remark:

- When $\phi(\cdot) = \frac{1}{2\gamma}\|\cdot\|^2$, $\gamma > 0$, F_ϕ is the Moreau envelope.
- When $f = h + P$, where $h \in C^2$ has Lipschitz gradient with modulus L, and $\phi(\cdot) = \frac{1}{2\gamma}\|\cdot\|^2 - h(\cdot)$, $\gamma \in (0, \frac{1}{L})$, F_ϕ is the forward-backward envelope. (Stella et al. '17)

Calculus of KL exponent IV

Theorem 5. (Yu, Li, P. '19)
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

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

If $\nabla G(\bar{x}) : \mathbb{Y} \to \mathbb{X}$ is injective and g_1 is a KL function with exponent α, then so is g.

Application VII

Corollary 6. (Yu, Li, P. '19)
Consider the function

$$f(X) := \frac{1}{2}\|\mathcal{A}X - b\|^2 + \delta_{\mathrm{rank}(\cdot)\leq r}(X)$$

for $X \in \mathbb{R}^{m\times n}$, where $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a linear map, $b \in \mathbb{R}^p$. Then f is KL with exponent $1 - \frac{1}{4\cdot 9^\kappa}$, where $\kappa = mn + m(m-r) + n(m-r) - 1$.

Application VII

Corollary 6. (Yu, Li, P. '19)
Consider the function

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

for $X \in \mathbb{R}^{m\times n}$, where $\mathcal{A} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a linear map, $b \in \mathbb{R}^p$. Then f is KL with exponent $1 - \frac{1}{4\cdot 9^\kappa}$, where $\kappa = mn + m(m-r) + n(m-r) - 1$.

Key proof idea:

$$f(X) = \inf_U \left\{ \frac{1}{2}\|\mathcal{A}X - b\|^2 + \frac{1}{2}\|U^T U - I_{m-r}\|_F^2 + \delta_{\mathfrak{D}}(X,U) + \delta_{\mathfrak{B}}(X,U) \right\},$$

where

$$\mathfrak{D} := \{(X,U) \in \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times(m-r)} :\ U^T X = 0\},$$

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

Conclusion

- KL exponent is an important quantity for determining the qualitative convergence behavior of first-order methods.
- We presented some rules for deducing KL exponents:
  - ⋆ Lagrangian relaxation.
  - ⋆ Min of finitely many functions.
  - ⋆ Inf-projection.

References:

- 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. 18:1199–1232, 2018.
- P. Yu, G. Li and T. K. Pong.
  Deducing Kurdyka-Łojasiewicz exponent via inf-projection.
  Available at https://arxiv.org/abs/1902.03635.

Thanks for coming! ☺