

Explicit estimation of KL exponent and linear convergence of 1st-order methods

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

ICCOPT 2016
August 2016
(Joint work with Guoyin Li)

Motivating applications

Sparse optimization problems:

- Logistic regression with ℓ_1 regularization:

$$\min_{x \in \mathbb{R}^n} \sum_{i=1}^m \log(1 + \exp(Ax)_i) + \mu \sum_{i=1}^{n-1} |x_i|.$$

- Logistic regression with sparsity constraint:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \sum_{i=1}^m \log(1 + \exp(Ax)_i) \\ \text{s.t.} \quad & \text{card}\{i : x_i \neq 0, 1 \leq i \leq n-1\} \leq r. \end{aligned}$$

- Can also consider least squares loss.

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 $k = 1, \dots$,

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

where

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

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 $\bar{x} \in \text{dom } \partial f$ if there exist $c, \nu, \epsilon > 0$ so that

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

whenever $x \in \text{dom } \partial f$, $\|x - \bar{x}\| \leq \epsilon$ 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 $\bar{x} \in \text{dom } \partial f$ if there exist $c, \nu, \epsilon > 0$ so that

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

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

- If f has the KL property at any $\bar{x} \in \text{dom } \partial f$ 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 $\bar{x} \in \text{dom } \partial f$ if there exist $c, \nu, \epsilon > 0$ so that

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

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

- If f has the KL property at any $\bar{x} \in \text{dom } \partial f$ 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. '15), etc., if f is replaced by a suitable potential function.

Existing results

For nonsmooth objectives:

- A convex piecewise linear-quadratic function is a KL function with exponent $\frac{1}{2}$. (Li '95, Bolte et al. '15)
- 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, Bolte et al. '15)

Existing results

For nonsmooth objectives:

- A convex piecewise linear-quadratic function is a KL function with exponent $\frac{1}{2}$. (Li '95, Bolte et al. '15)
- 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, Bolte et al. '15)
- 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)

Our strategy

Aim: Explicitly estimate the KL exponent of commonly used optimization models.

Strategy:

- Relate KL property to the Luo-Tseng error bound. (Luo, Tseng '92, '92, '93)
- Develop calculus rules on KL exponents: build new KL functions from old ones with known exponents.

Luo-Tseng error bound

Denote $\mathcal{X} := \{x : 0 \in \partial f(x)\}$, where $f = h + P$. Assume in addition that P is **convex**.

Definition: (Luo, Tseng '92, Tseng, Yun '09)

Suppose that $\mathcal{X} \neq \emptyset$. We say that the Luo-Tseng error bound holds if for any $\zeta \geq \inf f$, there exist $c, \epsilon > 0$ so that

$$\text{dist}(x, \mathcal{X}) \leq c \|\text{prox}_P(x - \nabla h(x)) - x\|$$

whenever $\|\text{prox}_P(x - \nabla h(x)) - x\| < \epsilon$ and $f(x) \leq \zeta$.

Luo-Tseng error bound

Denote $\mathcal{X} := \{x : 0 \in \partial f(x)\}$, where $f = h + P$. Assume in addition that P is **convex**.

Definition: (Luo, Tseng '92, Tseng, Yun '09)

Suppose that $\mathcal{X} \neq \emptyset$. We say that the Luo-Tseng error bound holds if for any $\zeta \geq \inf f$, there exist $c, \epsilon > 0$ so that

$$\text{dist}(x, \mathcal{X}) \leq c \|\text{prox}_P(x - \nabla h(x)) - x\|$$

whenever $\|\text{prox}_P(x - \nabla h(x)) - x\| < \epsilon$ and $f(x) \leq \zeta$.

Assumption 1: (Luo, Tseng '92, Tseng, Yun '09)

There exists $\delta > 0$ so that if $x, y \in \mathcal{X}$ and $\|x - y\| \leq \delta$, then $f(x) = f(y)$.

Luo-Tseng error bound

Examples: When $\mathcal{X} \neq \emptyset$ and $f = h + P$, Assumption 1 and the Luo-Tseng error bound hold for

- $h(x) = \ell(Ax)$ and P is proper polyhedral, where ℓ is strongly convex on any compact convex set and is twice continuously differentiable. (Luo, Tseng '92, Tseng, Yun '09)
- h is a quadratic (not necessarily convex) and P is proper polyhedral. (Luo, Tseng '92, Tseng, Yun '09)

Luo-Tseng error bound

Examples: When $\mathcal{X} \neq \emptyset$ and $f = h + P$, Assumption 1 and the Luo-Tseng error bound hold for

- $h(x) = \ell(Ax)$ and P is proper polyhedral, where ℓ is strongly convex on any compact convex set and is twice continuously differentiable. (Luo, Tseng '92, Tseng, Yun '09)
- h is a quadratic (not necessarily convex) and P is proper polyhedral. (Luo, Tseng '92, Tseng, Yun '09)

Theorem 1. (Li, P. '16)

Suppose that $\mathcal{X} \neq \emptyset$, and Assumption 1 and the Luo-Tseng error bound hold. Then f is a KL function with exponent $\frac{1}{2}$.

Calculus of KL exponent

Theorem 2. (Li, P. '16)

Suppose that g_i are KL functions with exponents α_i , $i = 1, \dots, m$. Suppose in addition that $g := \min_{1 \leq i \leq m} g_i$ is continuous on $\text{dom } \partial g$ and that $\text{dom } \partial g_i = \text{dom } g_i$ for all i . Then g is a KL function with exponent $\max\{\alpha_i : 1 \leq i \leq m\}$.

Calculus of KL exponent

Theorem 2. (Li, P. '16)

Suppose that g_i are KL functions with exponents α_i , $i = 1, \dots, m$. Suppose in addition that $g := \min_{1 \leq i \leq m} g_i$ is continuous on $\text{dom } \partial g$ and that $\text{dom } \partial g_i = \text{dom } g_i$ for all i . Then g is a KL function with exponent $\max\{\alpha_i : 1 \leq i \leq m\}$.

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

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

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

Application I

Corollary 1. (Li, P. '16)

Consider functions of the form

$$f(x) = \ell(Ax) + \min_{1 \leq i \leq m} P_i(x)$$

where ℓ is strongly convex on any compact convex set and is twice continuously differentiable, P_i are proper polyhedral functions. If f is continuous on $\text{dom } \partial f$, then f is a KL function with exponent $\frac{1}{2}$.

Application I

Corollary 1. (Li, P. '16)

Consider functions of the form

$$f(x) = \ell(Ax) + \min_{1 \leq i \leq m} P_i(x)$$

where ℓ is strongly convex on any compact convex set and is twice continuously differentiable, P_i are proper polyhedral functions. If f is continuous on $\text{dom } \partial f$, then f is a KL function with exponent $\frac{1}{2}$.

Example:

$$\begin{aligned} f(x) &= \ell(Ax) + \delta_{\|\cdot\|_0 \leq r}(x) \\ &= \ell(Ax) + \min_{I \in \mathcal{I}_{n-r}} \delta_{H_I}(x), \end{aligned}$$

where $\mathcal{I}_k := \{J \subseteq \{1, \dots, n\} : |J| = k\}$, $H_I := \{x : x_i = 0 \ \forall i \in I\}$.

Application II

Corollary 2. (Li, P. '16)

Consider functions of the form

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

where M_i are symmetric matrices, P_i are proper polyhedral functions. If f is continuous on $\text{dom } \partial f$, then f is a KL function with exponent $\frac{1}{2}$.

Application II

Corollary 2. (Li, P. '16)

Consider functions of the form

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

where M_i are symmetric matrices, P_i are proper polyhedral functions. If f is continuous on $\text{dom } \partial f$, 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

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

Conclusion

- The Luo-Tseng error bound together with an assumption on the separation of stationary values implies that the KL exponent is $\frac{1}{2}$.
- Based on this and some calculus rules for KL exponents, the KL exponent for a large class of convex/nonconvex optimization models is obtained, including
 - ★ logistic regression with ℓ_1 regularization/sparsity constraints;
 - ★ least squares problem with SCAD regularization.

Reference:

- G. Li and T. K. Pong.
Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods.
Available at <http://arxiv.org/abs/1602.02915>.

Thanks for coming! ☺