From error bounds to the complexity of first-order descent methods for convex functions

Nguyen Trong Phong-TSE

Joint work with Jérôme Bolte, Juan Peypouquet, Bruce Suter.

Toulouse, 23-25, March, 2016

Contents

Some notations

- Denote H is Hilbert space . $\langle , \rangle$ is a scalar product and $\|.\|$ is associated norm.
- ∂f : Subdifferential.
- prox_f : Proximal operator.
- $\text{dist}\,(x, S) = \min_{y \in S} \|x - y\|$.

Definition : Error Bound

f satisfies a local error bound if there exists a nondecreasing function $\varphi : [0,+\infty[\to [0,+\infty[$ and $r_0 > 0$ such that

$$\varphi([f(x)]_+) \geq \text{dist}(x, S), \forall x \in [0 \leq f \leq r_0],$$

where $S = \{x \in \mathbb{R}^n | f(x) \leq 0\}$, $[a]_+ = \max\{a, 0\}$.
φ : *residual function*

There are a lot of researchs on error bounds : A. Auslender, J.P Crouzeix, J.N Corvellec, A.J Hoffman,P. Tseng, Z.Q Luo, J.S Pang,...

Notation

Let $\eta > 0$ and set

$$\mathcal{K}(0,\eta) = \left\{ \varphi \in C^0[0,\eta) \cap C^1(0,\eta),\ \varphi(0) = 0,\ \varphi \text{ is concave, } \varphi' > 0 \right\}.$$

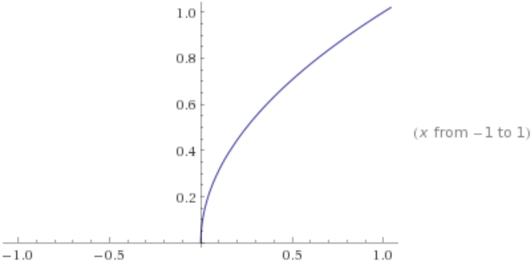

FIGURE: φ

Definition : Non-smooth KL property

$f : H \to (-\infty, \infty]$ has Kurdyka-Łojasiewicz (KL) property at $\bar{x}$ if there exists a neighbour $U(\bar{x})$, $\eta > 0$ and a function $\varphi \in \mathcal{K}(0, \eta)$ such that

$$\varphi'(f(x) - f(\bar{x}))\mathrm{dist}(0, \partial f(x)) \geq 1, \tag{1}$$

for all $x \in U(\bar{x}) \cap [f(\bar{x}) < f < f(\bar{x}) + \eta]$.
φ : Desingularizing function for f at $\bar{x}$.

Remark

- If $f(\bar{x}) = 0$ then (1) can be rewritten $\|\partial^0(\varphi \circ f)\| \geq 1$, where $\|\partial^0 f(x)\| = \inf \|\partial f(x)\|$.
- When $\varphi(s) = cs^{1-\theta}, \theta \in (0,1)$ then (1) is called Łojasiewicz inequality, $\|\partial^0 f(x)\|_-^\theta \geq c|f(x)|$.

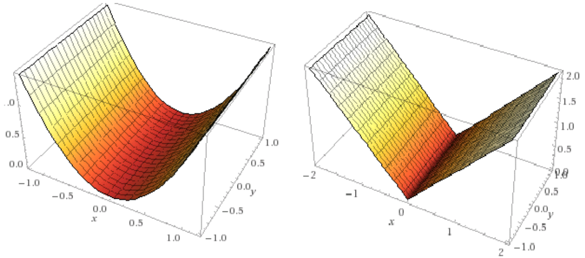

FIGURE: f and $\varphi \circ f$

The KL function class

- If f is analytic or smooth and semialgebraic, it satisfies the Lojasiewicz property around each point of $\mathbb{R}^n$, (S. Łojasiewicz, Hormander (1968), K. Kurdyka(1998)).
- $f : \mathbb{R}^n \to \mathbb{R} \cap \{+\infty\}$ lower semicontinuous and semi-algebraic (non-smooth), then f has the KL property around each point, (J. Bolte-A. Daniilidis-A. Lewis, 2006).

Applications of KL property for first order method

With the KL property, we can obtain the convergence of some methods (and its convergence rate). This can be seen in some references.

- Line-search, trust-region, (P.A. Absil-R. Mahony-B.Andrew 2005).
- Proximal method, (H.Attouch-J.Bolte, 2009).
- Forward-Backward method, (H. Attouch-J. Bolte-B.F. Svaiter 2014,).
- Proximal Alternating Linearized Minimization, (J.Bolte-M.Teboulle- S.Sabach, 2014).

KL has a lot of applications, however it is not easy to find the desingularizing φ, even the exponent θ in the Łojasiewicz inequality.

Theorem

Let $f : H \to]-\infty, +\infty]$ *be a proper, convex and lower-semicontinuous, with* $\min f = 0$*. Let* $r_0 > 0$, $\varphi \in \mathcal{K}(0, r_0)$, $c > 0$, $\rho > 0$ *and* $\bar{x} \in \operatorname{argmin} f$.

(i) *If* $\varphi'(f(x)) \|\partial^0 f(x)\| \geq 1$ *for all* $x \in [0 < f < r_0] \cap B(\bar{x}, \rho)$, *then*

$$\varphi(f(x)) \geq \operatorname{dist}(x, S), \forall x \in [0 < f < r_0] \cap B(\bar{x}, \rho).$$

(ii) *Conversely, if* $s\varphi'(s) \geq c\varphi(s)$ *for all* $s \in (0, r_0)$ *and* $\varphi(f(x)) \geq \operatorname{dist}(x, S)$ *for all* $x \in [0 < f < r_0] \cap B(\bar{x}, \rho)$, *then*

$$\varphi'(f(x)) \|\partial^0 f(x)\| \geq c, \forall x \in [0 < f < r_0] \cap B(\bar{x}, \rho).$$

Definition : Subgradient descent sequence

$(x_k)_{k\in\mathbb{N}}$ in H is said *subgradient descent sequence* for $f : H \to]-\infty, \infty]$ if $x_0 \in \operatorname{dom} f$ and there exist $a, b > 0$ such that :

(H1) (Sufficient decrease condition) For each $k \geq 1$,

$$f(x_k) + a\|x_k - x_{k-1}\|^2 \leq f(x_{k-1}).$$

(H2) (Relative error condition) For each $k \geq 1$, there is $\omega_k \in \partial f(x_k)$ such that

$$\|\omega_k\| \leq b\|x_k - x_{k-1}\|.$$

This definition encompasses many methods

- Projection gradient.
- Forward-Backward.
- Proximal alternating linearised minimization.

Notation

Fix f convexe, KL function (with φ is the desingalurizing function) and a descent method.
Set $\psi = (\varphi|_{[0,r_0]})^{-1} : [0, \alpha_0] \to [0, r_0]$ which ψ' is l-Lipschitz continuous (on $[0, \alpha_0]$) and $\psi'(0) = 0$.
Set $c = \frac{\sqrt{1+2l\,a\,b^{-2}}-1}{l}$,
($a > 0$, $b > 0$ are parameters of descent method).
Take $\alpha_0 = \varphi(f(x_0) - \min f)$, we define $(\alpha_k)_{k\in\mathbb{N}}$ by

$$\begin{aligned}\alpha_{k+1} &= \operatorname{argmin}\left\{\psi(u) + \frac{1}{2c}(u-\alpha_k)^2 : u \geq 0\right\}\\ &= \operatorname{prox}_{c\psi}(\alpha_k), \forall k \geq 0.\end{aligned}$$

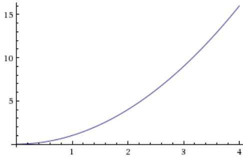

FIGURE: ψ

Main result

Theorem (Complexity of descent sequences for convex KL functions)

$f : H \to]-\infty, \infty]$ be a proper lower-semicontinuous convex function, which have the KL property on $[\min f < f < \min f + \eta]$, $\operatorname{argmin} f \neq \emptyset$. *$(x_k)_{k\in\mathbb{N}}$ be a subgradient descent sequence with $f(x_0) = r_0 \in (0, \eta)$. Then, x_k converges to some minimizer x^* and, moreover,*

$$f(x_k) - \min f \leq \psi(\alpha_k) \quad \forall k \geq 0,$$

$$\|x_k - x^*\| \leq \frac{b}{a}\alpha_k + \sqrt{\psi(\alpha_{k-1})}, \quad \forall k \geq 1.$$

Our work / our methodology

Take a first order method for a problem $\min f$ and set $r_0 = f(x_0)$

- Derive an error bound for the objective

$$f(x) - \min f \geq \psi(\mathrm{dist}\,(x, \mathrm{argmin}\, f))$$

  for all x such that $f(x) \leq r_0$.

- Study the *worst case* one dimensional method

$$\alpha_k = \mathrm{argmin}\left\{ c\psi(s) + \frac{1}{2}(s - \alpha_{k-1})^2 : s \geq 0 \right\},\ \alpha_0 = \varphi(f(x_0)),$$

  where c is a constant (easily) computed from the parameters of the first order method.

- Our complexity result asserts that

$$f(x_k) - \min f \leq \psi(\alpha_k) = \psi \circ \mathrm{prox}_{c\psi} \circ \ldots \circ \mathrm{prox}_{c\psi}(\varphi(f(x_0))).$$

Example : The l^1-regularized least squares problem

Problem

$$\min_{\mathbb{R}^n} \left\{ f(x) = \mu\|x\|_1 + \frac{1}{2}\|Ax - b\|_2^2 \right\},$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. We denote

- $\tilde{A} = [A, 0_{\mathbb{R}^{m\times 1}}] \in \mathbb{R}^{m\times(n+1)},\ \tilde{b} = (b_1, \ldots, b_m, 0) \in \mathbb{R}^{m+1}$.
- $\tilde{\mu} = (0, \ldots, 0, \mu) \in \mathbb{R}^{n+1},\ \tilde{x} = (x, y) \in \mathbb{R}^{n+1},\ \tilde{R} = (0, \ldots, 0, R) \in \mathbb{R}^{n+1}$.
- $M = \begin{bmatrix} E & -1_{\mathbb{R}^{2^n\times 1}} \\ 0_{\mathbb{R}^{1\times n}} & 1 \end{bmatrix}$ is a matrix of size $(2^n + 1) \times (n + 1)$, where E is a matrix of size $2^n \times n$ whose rows are all possible distinct vectors of size n of the form $e_i = (\pm 1, \ldots, \pm 1)$ for all $i = 1, \ldots, 2^n$. The order of the e_i being arbitrary.

Calculating error bound

Using the work of Beck-Shtern, 2015 :

Lemma

Fix $R > \frac{\|b\|^2}{2\mu}$. Then, for all $x \in \mathbb{R}^n$ such that $\|x\|_1 \leq R$, we have

$$f(x) - f(x^*) \geq \frac{\gamma_R}{2} \operatorname{dist}^2(x, S),$$

where $\gamma_R^{-1} = \nu^2 \left(1 + \frac{\sqrt{5}}{2}\mu R + (R\|A\| + \|b\|)\,(4R\|A\| + \|b\|)\right)$, and ν is the Hoffman constant associated with $(M, [\tilde{A}^T, \tilde{\mu}^T]^T)$.
Therefore, f is a KL function on the ℓ^1 ball of radius R and admits $\varphi(s) = \sqrt{2\gamma_R^{-1} s}$ as desingularizing function.

Complexity

Take $x_0 \in \mathbb{R}^n$, the Forward-Backward method applied to f is

$$x_{k+1} = \operatorname{prox}_{\lambda_k \mu \|\cdot\|_1} \left(x_k - \lambda_k (A^T A x_k - A^T b) \right) \quad \text{for } k \geq 0,$$

where $0 < \lambda^- \leq \lambda_k \leq \lambda^+ < 2/L$, and $L = \|A^T A\|$. Set

$$\zeta = \frac{\sqrt{1 + \gamma \left(\frac{2}{\lambda^+} - L \right) \left(\frac{1}{\lambda^-} + L \right)^{-2}} - 1}{\gamma}.$$

Complexity and convergence rates

$(x_k)_{k \in \mathbb{N}}$ converges to a minimizer x^* of f and satisfies,

$$f(x_k) - \min f \leq \frac{1}{(1 + \gamma\zeta)^{2k}} (f(x_0) - \min f), \forall k \geq 0,$$

$$\|x^* - x_k\| \leq \frac{\sqrt{2(f(x_0) - \min f)}}{\sqrt{\gamma}\,(1 + \gamma\zeta)^{k-1}} \left(\sqrt{\frac{\gamma}{2}} + \frac{2(1 + L\lambda^-)}{(2 - L\lambda^+)(1 + \zeta\gamma)} \right), \forall k \geq 1.$$

More details :
Jérôme Bolte, Trong Phong Nguyen, Juan Peypouquet, Bruce Suter : From error bounds to the complexity of first-order descent methods for convex functions, *http ://arxiv.org/pdf/1510.08234.pdf*

THANK YOU FOR YOUR ATTENTION