Convergence of Cubic Regularization for Nonconvex Optimization under Łojasiewicz Property

Yi Zhou[*] Zhe Wang[†] Yingbin Liang[†]

∗ Department of ECE, Duke University
† Department of ECE, The Ohio State University

Cubic-regularization (CR)

$$\min_{x \in \mathbb{R}^d} f(x)$$

$$\text{(CR):}\;\; x_{k+1} \in \text{argmin}_y \langle y - x_k, \nabla f(x_k) \rangle + \frac{1}{2}(y - x_k)^\top \nabla^2 f(x_k)(y - x_k) + \frac{M}{6}\|y - x_k\|^3$$

- Converge to 2nd-order stationary point (Nesterov'06)

$$(2^{\text{nd}} - \text{order stationary}) \quad \nabla f(x) = 0, \nabla^2 f(x) \succcurlyeq 0.$$

- Escape strict-saddle points

Motivation and Contribution

- General nonconvex optimization
  - global sublinear convergence (Nesterov'06)
- Nonconvex + local geometry
  - gradient dominance (Nesterov'06)
    - super-linear convergence
  - error bound (Yue'18)
    - quadratic convergence
  - limited function class

- Our contributions
  - general Łojasiewicz property

Lojasiewicz Property

Definition (Lojasiewicz Property) Let f takes a constant value f^* on a compact set Ω. There exists $\epsilon, \lambda > 0$ such that for all $x \in \{z \in \Re^d : \mathrm{dist}_\Omega(z) < \epsilon, f^* < f(z) < f^* + \lambda\}, one\ has$

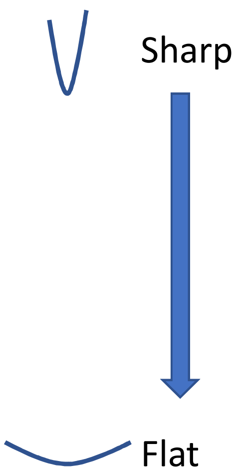
$$f(x) - f^* \leq \|\nabla f(x)\|^\theta,$$

where $\theta \in (1, +\infty)$ is the Lojasiewicz exponent.

- Satisfied by large function class:
  - analytic function, polynomials, exp-log functions, etc
  - ML examples: Lasso, phase retrieval, blind deconvolution, etc.

Convergence to 2nd-order Stationary Point

$$\mu(x) := \max\left\{\sqrt{\|\nabla f(x)\|}, -\lambda_{\min}\left(\nabla^2 f(x)\right)\right\}$$



Lojasiewicz exponent $\boldsymbol{\theta}$	Convergence rate
$\theta = +\infty$	$\mu(x_{k_0}) = 0$ finite-step
$\theta \in \left(\frac{3}{2}, +\infty\right)$	$\mu(x_k) \le \Theta\left(\exp -(2(\theta-1))^{k-k_0}\right)$ super-linear
$\theta = \frac{3}{2}$	$\mu(x_k) \le \Theta(\exp(-(k-k_0)))$ linear
$\theta \in \left(1, \frac{3}{2}\right)$	$\mu(x_k) \le \Theta\left((k-k_0)^{-\frac{2(\theta-1)}{3-2\theta}}\right)$ sub-linear

Convergence of Function Value

Lojasiewicz exponent $\boldsymbol{\theta}$	Convergence rate
$\theta = +\infty$	$f\left(x_{k_0}\right) - f^* = 0$
$\theta \in \left(\frac{3}{2}, +\infty\right)$	$f(x_k) - f^* \leq \Theta\left(\exp - \left(\frac{2\theta}{3}\right)^{k-k_0}\right)$
$\theta = \frac{3}{2}$	$f(x_k) - f^* \leq \Theta(\exp(-(k - k_0)))$
$\theta \in \left(1, \frac{3}{2}\right)$	$f(x_k) - f^* \leq \Theta\left((k - k_0)^{-\frac{2\theta}{3-2\theta}}\right)$

Convergence of Variable Sequence

Theorem Assume f satisfies the Lojasiewicz property. Then, the sequence generated by CR is absolutely-summable as

$$\Sigma_{k=0}^{\infty}\|x_{k+1} - x_k\| < +\infty.$$

- Implies Cauchy-convergent

- (Nesterov'06): cubic-summable

$$\Sigma_{k=0}^{\infty}\|x_{k+1} - x_k\|^{\mathbf{3}} < +\infty$$

Convergence of Variable Sequence

Lojasiewicz exponent $\boldsymbol{\theta}$	Convergence rate
$\theta = +\infty$	$x_{k_0} - x^* = 0$
$\theta \in \left(\frac{3}{2}, +\infty\right)$	$\|x_k - x^*\| \leq \Theta\left(\exp - \left(\frac{2(\theta-1)}{3} + \frac{2}{3}\right)^{k-k_0}\right)$
$\theta = \frac{3}{2}$	$\|x_k - x^*\| \leq \Theta(\exp(-(k - k_0)))$
$\theta \in \left(1, \frac{3}{2}\right)$	$\|x_k - x^*\| \leq \Theta\left((k - k_0)^{-\frac{2(\theta-1)}{3-2\theta}}\right)$

Comparison with First-order Algorithm

Lojasiewicz exponent $\boldsymbol{\theta}$	Gradient descent	Cubic-regularization
$\theta = +\infty$	finite-step	finite-step
$\theta \in (2, +\infty)$	linear	super-linear
$\theta \in [\frac{3}{2}, 2)$	sub-linear	super-linear
$\theta \in \left(1, \frac{3}{2}\right)$	sub-linear $\Theta(k^{-\frac{\theta-1}{2-\theta}})$	sub-linear $\Theta(k^{-\frac{\theta-1}{1.5-\theta}})$

Come to our poster

Thursday 05:00 PM

Room 210 & 230 AB #4

Thank You!