CONVERGENCE OF DESCENT METHODS UNDER POLYAK-KURDYKA-ŁOJASIEWICZ PROPERTIES

BORIS MORDUKHOVICH

Wayne State University, USA

talk given at the ICOMP Conference

based on joint work with G. C. Bento, T. S. Mota and Yu. Nesterov

Innopolis, Russia, October 2024

This talk presents comprehensive convergence analysis of a generic class of descent methods in nonsmooth and nonconvex optimization under several versions of the Polyak-Kurdyka-Łojasiewicz (PKL) property. Along other results, we prove the finite termination of the generic algorithm under the PKL property with lower exponents $0 < q < 1/2$. Specifications are given to convergence rates of some particular algorithms including inexact reduced gradient methods and the boosted algorithm in DC programming. It is revealed e.g., that the lower exponent PKL property in the DC framework is incompatible with the gradient Lipschitz continuity for the plus function around a local minimizer. On the other hand, we show that the above inconsistency observation fails if the Lipschitz continuity is replaced by merely the gradient continuity.

DEFINITION. Let $f : \mathbb{R}^n \to \overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ be lower semicontinuous (l.s.c.) function with the domain $\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < \infty\}$. We say that the function f satisfies:

(i) The basic Polyak-Kurdyka-Łojasiewicz property at $\overline{x} \in \operatorname{dom} f$ if there exist a number $\eta \in (0, \infty)$, a neighborhood U of $\overline{x}$, and a concave continuous function $\varphi : [0, \eta] \to [0, \infty)$, called the desingularizing function, such that

$$\varphi(0) = 0, \quad \varphi \in C^1(0, \eta), \quad \varphi'(s) > 0 \quad \forall s \in (0, \eta),$$

$$\varphi'\big(f(x) - f(\overline{x})\big)\operatorname{dist}\big(0, \partial f(x)\big) \geq 1 \quad \forall x \in U \cap [f(\overline{x}) < f(x) < f(\overline{x}) + \eta],$$

where $\partial f(\overline{x})$ stands for the limiting subdifferential [M76] of f at $\overline{x}$

$$\partial f(\overline{x}) := \Big\{ v \in \mathbb{R}^n \;\Big|\; \exists x_k \to \overline{x},\ v_k \to v \ \text{ with } \ f(x_k) \to f(x),\ \limsup_{u \to x_k} \frac{f(u) - f(x_k) - \langle v_k, u - x_k \rangle}{\|u - x_k\|} \geq 0,\ k \in \mathbb{N} \Big\}.$$

(ii) The symmetric PKL property at $\overline{x}$ if f is continuous around $\overline{x}$ and ∂f is replaced by the symmetric subdifferential of f at $\overline{x}$ defined by

$$\partial^0 f(\overline{x}) := \partial f(\overline{x}) \cup \big(-\partial(-f)(\overline{x})\big).$$

(iii) The strong PKL property holds if f is Lipschitz continuous around $\overline{x}$ and ∂ is replaced in by the convexified/Clarke subdifferential of f

$$\overline{\partial} f(\overline{x}) := \mathrm{co}\, \partial f(\overline{x})$$

where "co" stands for the convex hull of the set in question.

(iv) The exponent PKL properties in (i)–(iii) selects $\varphi(t) = Mt^{1-q}$ with $M > 0$ and $q \in [0, 1)$. We refer to the case where $q \in (0, 1/2)$ as the PKL property with lower exponents.

The PKL property was originated independently by Łojasiewicz [L63] as

$$\|\nabla f(x)\| \geq b\,|f(x) - f(\overline{x})|^q, \quad b := 1/M(1-q),$$

in the general theory of real analytic functions, and by Polyak [P63] when $q = 1/2$ for functions of class $\mathcal{C}^{1,1}$ (i.e., $\mathcal{C}^1$ functions with Lipschitzian gradients) who used it to prove linear convergent of the gradient descent method. This property is known in optimization as the Polyak-Łojasiewicz inequality. Kurdyka (1998) proposed extensions to general algebraic-geometric structures. Nonsmooth versions of PKL in (i,iii,iv) were suggested by Attouch, Bolté et al.[AB09,ABS13,BDL06]. The symmetric version in (ii) is new.

The required tools of variational analysis and generalized differentiation can be found in [M06,RW98] and the references therein.

In [ABS13], Attouch, Bolté and Svaiter consider minimizing l.s.c. functions $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ by using a generic class of descent methods satisfying the conditions:

($\mathcal{H}1$) *Sufficient decrease condition*: for each $k \in \mathbb{N}$ we have

$$f(x^{k+1}) + a\|x^{k+1} - x^k\|^2 \le f(x^k).$$

($\mathcal{H}2$) *Relative error condition*: for each $k \in \mathbb{N}$ we have

$$\exists\, w^{k+1} \in \partial f(x^{k+1}) \text{ with } \|w^{k+1}\| \le b\|x^{k+1} - x^k\|$$

with $a, b > 0$. It is shown in [ABS13] that a great variety of important algorithms of optimization satisfies conditions ($\mathcal{H}1$) and ($\mathcal{H}2$). Employing the basic PKL property, a generic convergence analysis with arriving at the limiting/M-stationary point $0 \in \partial f(\overline{x})$ is developed in [ABS13] for the class of descent algorithms satisfying ($\mathcal{H}1$) and ($\mathcal{H}2$) while without establishing any convergence rate.

There exist remarkable descent methods for which the relative error condition $(\mathcal{H}2)$ fails. E.g., it was observed by Aragón-Artacho and Vuong [AV20] for the Boosted DCA (BDCA) in DC programming. They provided convergence analysis, with convergence to C-stationary point $0 \in \overline{\partial} f(\overline{x})$, by the replacement of $(\mathcal{H}2)$ with

$$\exists \;\; w^k \in \overline{\partial} f(x^k) \;\; \text{with} \;\; \|w^k\| \leq b\|x^{k+1} - x^k\|$$

via the convexified subdifferentialof the difference $f = g - h$ of strongly convex functions under the strong PKL in (iii).

We consider a generic class of descent algorithms satisfying sufficient decrease condition $(\mathcal{H}1)$ and with replacing $(\mathcal{H}2)$ by the new

$(\mathcal{H}3)$ Modified error condition: for each $k \in \mathbb{N}$

$$\exists \;\; w^k \in \partial f(x^k) \;\; \text{with} \;\; \|w^k\| \leq b\|x^{k+1} - x^k\|$$

expressed in terms of the limiting subdifferential. We also impose the following technical assumption:

$(\mathcal{H}4)$ *Continuity condition*: There exist a subsequence $\{x^{k_j}\}_{j \in \mathbb{N}}$ and a point $\overline{x}$ such that

$$x^{k_j} \longrightarrow \overline{x} \;\; \text{and} \;\; f(x^{k_j}) \longrightarrow f(\overline{x}) \;\; \text{as} \;\; j \longrightarrow \infty.$$

THEOREM Let $f : \mathbb{R}^n \longrightarrow \overline{\mathbb{R}}$ be a proper l.s.c. function bounded from below, and let the sequence $\{x^k\}$ be constructed by a generic algorithm satisfying $(\mathcal{H}1)$, $(\mathcal{H}3)$, $(\mathcal{H}4)$. If the basic PKL property holds at some accumulation point $\overline{x}$ of $\{x^k\}$, then we have that

$$\sum_{k=0}^{\infty} \|x^k - x^{k+1}\| < \infty$$

and that the sequence $\{x^k\}$ converges to $\overline{x}$ as $k \to \infty$. Moreover, $\overline{x}$ is an *M*-stationary point of the function f.

THEOREM In the setting of the previous theorem, assume that the exponent PKL property of f holds at $\bar{x}$ with $\varphi(t) = Mt^{1-q}$ for some $M > 0$ and $q \in [0, 1)$. The following convergence rates are guaranteed for the generic iterative sequences:

(i) If $q \in [0, \frac{1}{2})$, then $\{x^k\}$ and $\{f(x^k)\}$ converge in a finite number of steps to $\bar{x}$ and $f(\bar{x})$, respectively.

(ii) If $q = \frac{1}{2}$, then the sequences $\{x^k\}$ and $\{f(x^k)\}$ converge linearly to $\overline{x}$ and $f(\bar{x})$, respectively.

(iii) If $q \in (1/2, 1)$, then there exists $\sigma > 0$ such that

$$\left\|x^k - \overline{x}\right\| \leq \sigma k^{-\frac{1-q}{2q-1}} \quad \text{for all large } \; k \in \mathbb{N}.$$

INEXACT GRADIENT REDUCED METHODS

A class of inexact reduced graduate (IRG) methods with various stepsize selections have been recently proposed and developed by Khanh, Mordukhovich and Tran [KMT24] for problems of smooth nonconvex optimization. The general model involves the iterative procedure

$$f(x^k) - f(x^{k+1}) \geq \frac{\beta}{t_k}\|x^{k+1} - x^k\|^2, \quad \|\nabla\phi(x^k)\| \leq \frac{c}{t_k}\|x^{k+1} - x^k\|$$

where $\{t_k\} \subset \mathbb{R}_+$ and $\beta, c > 0$. The convergence analysis of IRG in [KMP24] is based on the following observations:

• *Primary descent condition*: There exists $\sigma > 0$ such that

$$f(x^k) - f(x^{k+1}) \geq \sigma\|\nabla f(x^k)\| \cdot \|x^{k+1} - x^k\|$$

•*Complementary descent condition*: We have the implication

$$\left[f(x^{k+1}) = f(x^k)\right] \Longrightarrow \left[x^{k+1} = x^k\right]$$

The results presented above allow us to significantly improve those in [KMT24] and previous publications.

Consider the DC program

$$(\mathcal{P}) \qquad \min_{x\in\mathbb{R}^n} f(x) := g(x) - h(x)$$

where both $g, h : \mathbb{R}^n \to \overline{\mathbb{R}}$ are convex with g being $\mathcal{C}^1$-smooth

Algorithm (BDCA [AV20])

1. Fix $\alpha > 0, \overline{\lambda} > 0$, $\beta \in (0,1)$. Let x_0 be an initial point, $k := 0$.
2. Select $u_k \in \partial h(x^k)$ and solve the convex subproblem

$$(\mathcal{P}k') \qquad \min_{x\in\mathbb{R}^n} g(x) - \langle u_k, x\rangle$$

to obtain the unique solution y^k.

3. Let $d^k := y^k - x^k$. If $d^k = 0$, **STOP** and **RETURN** x^k. Otherwise, go to Step 4.
4. Choose any $\overline{\lambda}_k \geq 0$. Set $\lambda_k := \overline{\lambda}_k$. **WHILE** $f(y^k + \lambda_k d^k) > f(y^k) - \alpha\lambda_k^2\|d^k\|^2$ **DO** $\lambda_k := \beta\lambda_k$.
5. Let $x^{k+1} := y^k + \lambda_k d^k$. If $x^{k+1} = x^k$, **STOP** and **RETURN** x^k. Otherwise, set $k := k+1$ and go to Step 2.

Here is the application of our general results to the convergence and convergence rates of BDCA for continuous functions with the usage of symmetric PKL (ii) in terms of the symmetric subdifferential.

THEOREM. Consider problem $(\mathcal{P})$, where f is continuous . Let $\{x^k\}$ be generated by BDCA, and let ∇g be L-Lipschitz continuous around an accumulation point $\bar{x}$ of $\{x^k\}$. Then we have

$$f(x^{k+1}) \leq f(x^k) - \frac{\alpha\lambda_k^2 + \rho}{(1+\lambda_k)^2}\|x^{k+1} - x^k\|^2,$$

$$\exists\, w^k \in \partial^0 f(x^k) \ \text{ such that } \ \|w^k\| \leq L\|x^{k+1} - x^k\|$$

and $\{x^k\}$ converges to $\bar{x}$ satisfying $0 \in \partial^0 f(\bar{x})$. If f has the exponent PKL property with $\varphi(t) = Mt^{1-q}$ for some $M > 0$ and $q \in [0, 1)$, then the convergence rates of $x^k \to \bar{x}$ are as established above.

The following result shows the inconsistency of the lower exponent PKL property with Lipschitz continuity of gradients.

THEOREM. Let $\bar{x} \in \mathrm{int}(\mathrm{dom}\, h)$ be a local minimizer of the problem

$$\min_{x \in \mathbb{R}^n} f(x) := g(x) - h(x)$$

where g is of class $\mathcal{C}^{1,1}$ around $\bar{x}$, and where h is convex. Then the exponent PKL property of f at $\bar{x}$ fails whenever $q \in (0, 1/2)$.

Consider the function $f(x) := |x|^{3/2}$ having its global minimum at $\bar{x} = 0$. It can be directly checked that the derivative of f is not locally Lipschitzian around $\bar{x}$, while the lower exponent PKL property holds with $\varphi(t) = t^{1-q}$ and $q = 1/3$. This shows that the Lipschitz continuity of the gradient is essential for the inconsistency result.

[AV20] F. J. Aragón-Artacho and P. T. Vuong, The boosted difference of convex functions algorithm for nonsmooth functions, *SIAM J. Optim.* **30** (2020), 980–1006.

[AB09] H. Attouch and J. Bolté, On the convergence of the proximal algorithm for nonsmooth functions involving analytic features, *Math. Program.* **116** (2009), 5–16 .

[ABS13] H. Attouch, J. Bolté and B. F. Svaiter, Convergence of descent methods for semialgebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized Gauss–Seidel methods, *Math. Program.* **137** (2013), 91–129.

[BMMN24] G. Bento, B. S. Mordukhovich, T. Mota and Yu. Nesterov, Convergence of descent methods under Polyak-Kurdyka-Łojasiewicz properties (2024); arxiv:2407.00812.

[BDL06] J. Bolté, A. Daniilidis and A. S. Lewis, The Łojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems, *SIAM J. Optim.* **17** (2006), 1205—1223.

[KMT24] P. D. Khanh, B. S. Mordukhovich and D. B. Tran, Inexact reduced gradient methods in nonconvex optimization, *J. Optim. Theory Appl.* DOI: 10.1007/s10957-023-02319-9 (2024).

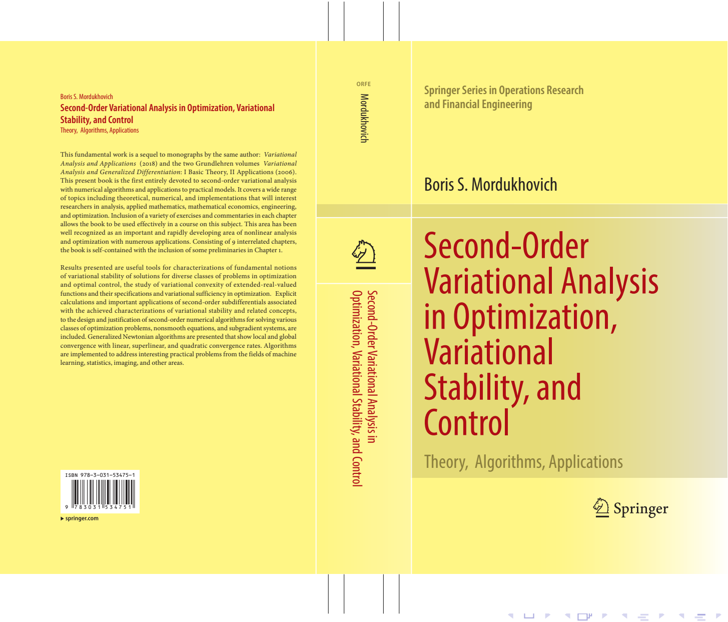
[L63] S. Łojasiewicz, Une propriété topologique des sous-ensembles analytiques réels, *Coll. du CNRS, Les équations aux dérivées partielles*, pp. 87–89 (1963).

[M76] B. S. Mordukhovich, Maximum principle in problems of time-optimal control with nonsmooth constraints, *J. Appl. Math. Mech.* **40** (1976), 960–969.

[M06] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation I: Basic Theory*, Springer, Berlin, 2006.

[P63] B. T. Polyak, Gradient methods for the minimization of functionals, *USSR Comput. Math. Math. Phys.* **3** (1963), 864–878.

[RW98] R. T. Rockafellar and R. J-B Wets, *Variational Analysis*, Springer, Berlin, 1998.


Boris S. Mordukhovich
Second-Order Variational Analysis in Optimization, Variational Stability, and Control
Theory, Algorithms, Applications
This fundamental work is a sequel to monographs by the same author: Variational Analysis and Applications (2018) and the two Grundlehren volumes Variational Analysis and Generalized Differentiation: I Basic Theory, II Applications (2006). This present book is the first entirely devoted to second-order variational analysis with numerical algorithms and applications to practical models. It covers a wide range of topics including theoretical, numerical, and implementations that will interest researchers in analysis, applied mathematics, mathematical economics, engineering, and optimization. Inclusion of a variety of exercises and commentaries in each chapter allows the book to be used effectively in a course on this subject. This area has been well recognized as an important and rapidly developing area of nonlinear analysis and optimization with numerous applications. Consisting of 9 interrelated chapters, the book is self-contained with the inclusion of some preliminaries in Chapter 1.
Results presented are useful tools for characterizations of fundamental notions of variational stability of solutions for diverse classes of problems in optimization and optimal control, the study of variational convexity of extended-real-valued functions and their specifications and variational sufficiency in optimization. Explicit calculations and important applications of second-order subdifferentials associated with the achieved characterizations of variational stability and related concepts, to the design and justification of second-order numerical algorithms for solving various classes of optimization problems, nonsmooth equations, and subgradient systems, are included. Generalized Newtonian algorithms are presented that show local and global convergence with linear, superlinear, and quadratic convergence rates. Algorithms are implemented to address interesting practical problems from the fields of machine learning, statistics, imaging, and other areas.
ISBN 978-3-031-53475-1
springer.com
ORFE
Mordukhovich
Second-Order Variational Analysis in Optimization, Variational Stability, and Control
Springer Series in Operations Research and Financial Engineering
Boris S. Mordukhovich
Second-Order Variational Analysis in Optimization, Variational Stability, and Control
Theory, Algorithms, Applications
Springer