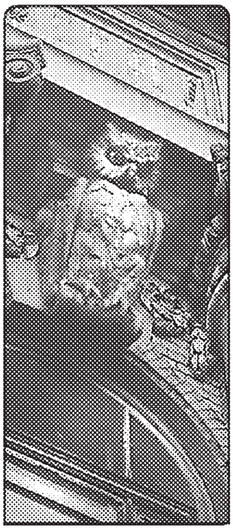

Научный отдел

МАТЕМАТИКА

Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2023. Т. 23, вып. 1. С. 4–10
Izvestiya of Saratov University. Mathematics. Mechanics. Informatics, 2023, vol. 23, iss. 1, pp. 4–10
mmi.sgu.ru
https://doi.org/10.18500/1816-9791-2023-23-1-4-10
EDN: ZSKZLA

Article

The Lezanski – Polyak – Lojasiewicz inequality and the convergence of the gradient projection algorithm

M. V. Balashov

V. A. Trapeznikov Institute of Control Sciences, 65 Profsoyuznaya St., Moscow 117977, Russia

Maxim V. Balashov, balashov73@mail.ru, https://orcid.org/0000-0002-5414-2149, AuthorID: 12909

Abstract. We consider the Lezanski – Polyak – Lojasiewicz inequality for a real-analytic function on a real-analytic compact manifold without boundary in finite-dimensional Euclidean space. This inequality emerged in 1963 independently in works of three authors: Lezanski and Lojasiewicz from Poland and Polyak from the USSR. The inequality is appeared to be a very useful tool in the convergence analysis of the gradient methods, firstly in unconstrained optimization and during the past few decades in problems of constrained optimization. Basically, it is applied for a smooth in a certain sense function on a smooth in a certain sense manifold. We propose the derivation of the inequality from the error bound condition of the power type on a compact real-analytic manifold. As an application, we prove the convergence of the gradient projection algorithm of a real analytic function on a real analytic compact manifold without boundary. Unlike known results, our proof gives explicit dependence of the error via parameters of the problem: the power in the error bound condition and the constant of proximal smoothness first of all. Here we significantly use a technical fact that a smooth compact manifold without boundary is a proximally smooth set.
Keywords: Lezanski – Polyak – Lojasiewicz inequality, error bound condition, proximal smoothness, gradient projection algorithm, real-analytic function
Acknowledgements: This work was supported by the Russian Science Foundation (project No. 22-11-00042).

For citation: Balashov M. V. The Lezanski – Polyak – Lojasiewicz inequality and the convergence of the gradient projection algorithm. *Izvestiya of Saratov University. Mathematics. Mechanics. Informatics*, 2023, vol. 23, iss. 1, pp. 4–10. https://doi.org/10.18500/1816-9791-2023-23-1-4-10, EDN: ZSKZLA

Научная статья
УДК 519.853.4

Неравенство Лежанского – Поляка – Лоясевича и сходимость метода проекции градиента

М. В. Балашов

Институт проблем управления РАН им. В. А. Трапезникова, Россия, 117977, г. Москва, ул. Профсоюзная, д. 65

Балашов Максим Викторович, доктор физико-математических наук, ведущий научный сотрудник, balashov73@mail.ru, https://orcid.org/0000-0002-5414-2149, AuthorID: 12909

Аннотация. Рассматривается неравенство Лежанского – Поляка – Лоясевича для вещественно-аналитической функции на вещественно-аналитическом компактном многообразии без края в конечномерном евклидовом пространстве. Это неравенство возникло независимо в 1963 г. в работах трех авторов: Лежанского и Лоясевича из Польши и Поляка из СССР. Неравенство оказалось очень полезным инструментом для исследования сходимости градиентных методов, первоначально в безусловной оптимизации, а в течение последних нескольких десятилетий и в задачах условной оптимизации. Оно применяется, главным образом, для гладких в определенном смысле функций на гладких в определенном смысле многообразиях. Мы предлагаем вывод неравенства из условия ограничения ошибки степенного типа на компактном вещественно-аналитическом многообразии. В качестве приложения мы доказываем сходимость метода проекции градиента вещественно-аналитической функции на вещественно аналитическом многообразии без края. В отличие от известных результатов, наше доказательство дает явную зависимость погрешности через параметры задачи: в первую очередь, через показатель в условии ограничения ошибки и константу проксимальной гладкости. При этом мы существенно используем технический факт, что гладкое компактное многообразие без края есть проксимально гладкое множество.
Ключевые слова: неравенство Лежанского – Поляка – Лоясевича, условие ограничения ошибки, проксимальная гладкость, метод проекции градиента, вещественно-аналитическая функция
Благодарности: Работа выполнена при поддержке Российского научного фонда (проект № 22-11-00042).
Для цитирования: *Balashov M. V.* The Lezanski – Polyak – Lojasiewicz inequality and the convergence of the gradient projection algorithm [*Балашов М. В.* Неравенство Лежанского – Поляка – Лоясевича и сходимость метода проекции градиента] // Известия Саратовского университета. Новая серия. Серия: Математика. Механика. Информатика. 2023. Т. 23, вып. 1. С. 4–10. https://doi.org/10.18500/1816-9791-2023-23-1-4-10, EDN: ZSKZLA

Introduction

Let $\mathbb{R}^n$ be n-dimensional Euclidean space with the inner product $(\cdot,\cdot)$ and the norm $\|\cdot\|^2 = (\cdot,\cdot)$. Define for $r > 0$ and $a \in \mathbb{R}^n$ the ball $B_r(a) = \{x \in \mathbb{R}^n \ : \ \|x - a\| \leqslant r\}$.

Recall that $S \subset \mathbb{R}^n$ is C^k, m-dimensional *manifold without boundary,* $1 \leqslant m \leqslant n-1$, if there exists an atlas $\{(U_i, \varphi_i)\}_{i\in I}$, $\varphi_i \in C^k$, $U_i \subset \mathbb{R}^n$, with $S = \bigcup\limits_{i\in I} U_i$, U_i is an open subset of S and $U_i = \varphi_i^{-1}(V_i)$ for an open subset $V_i \subset \mathbb{R}^m$ for all i.

The smoothness of the manifold S means that for any point $x \in S$ there exist a tangent subspace T_x to S at the point $x \in S$ and a tangent plane $x + T_x$. It should be noted that T_x has dimension m for any $x \in S$. See [1] for details. We denote by $P_A x$ the metric projection of a point $x \in \mathbb{R}^n$ onto a closed set $A \subset \mathbb{R}^n$.

A function is called *real-analytic* if it is locally can be represented as a convergent power series. The sum, product, and composition of real-analytic functions are also real-analytic. Components of the Frechet gradient of a real-analytic function are real-analytic too. A manifold is called real-analytic if functions φ_i are real-analytic.

Denote for $A \subset \mathbb{R}^n$ and $x \in \mathbb{R}^n$ the distance $\varrho_A(x) = \varrho(x, A) = \inf\limits_{a\in A} \|x - a\|$. In his famous work [2], Lojasiewicz proved that for a real-analytic function $f : \mathbb{R}^n \supset U \to \mathbb{R}$, U is an open set, and for any compact set $\mathcal{K} \subset U$ there exist $\alpha > 0$ and $C > 0$ with

$$\varrho(x, \{y \in U \mid f(y) = 0\}) \leqslant C|f(x)|^\alpha \qquad \forall\ x \in \mathcal{K}. \tag{1}$$

Note that the power $\alpha \in (0, 1]$ depends on $\mathcal{K}$. The value of α is usually hard to estimate a priori.

Consider a problem $\min\limits_{x\in\mathbb{R}^n} f(x)$ with a differentiable function. Suppose that $\Omega = \operatorname{Arg}\min\limits_{x\in\mathbb{R}^n} f(x) \neq \varnothing$. Let $f_* = f(\Omega)$. The *Lezansli – Polyak – Lojasiewicz (further LPL)* inequality for this problem means that there exist $\nu > 0$ and $\beta \in (0, 2]$ with

$$\|f'(x)\|^\beta \geqslant \nu(f(x) - f_*) \quad \forall\ x \in \mathbb{R}^n.$$

This inequality and some more general variants were considered independently in 1963 by Lezanski, Polyak and Lojasiewicz, see [3] for details. Note that under the above assumptions the LPL inequality may or may not be true.

In the case $\beta = 2$ the LPL condition provides the linear rate of convergence for the gradient descent method for a Lipschitz differentiable function [4].

Consider the case of constrained optimization

$$\min_{x\in S} f(x) \tag{2}$$

with a real-analytic function f and a real-analytic compact manifold without boundary S.

The standard *gradient projection algorithm (further GPA)* for problem (2) is the next iteration process

$$x_0 \in S, \quad x_{k+1} = P_S(x_k - tP_{T_{x_k}} f'(x_k)), \quad t > 0. \tag{3}$$

It is known [5, Proposition 2.2] that the LPL inequality for problem (2) has the next local form $\forall\ x \in S$

$$\exists\ U \ni x (U - \text{open subset of } S)\ \exists\ \beta \in (0, 2],\ \nu > 0\ \forall\ y \in U\ \nu|f(y) - f(x)| \leqslant \|P_{T_y} f'(y)\|^\beta.$$

On the basis of the last property, the convergence of the gradient projection method for (2) was proved [5,6]. The main disadvantage of mentioned works is that in estimates of the error there are fundamentally unknown constants. Thus these results are "not constructive".

We plan to deduce the LPL condition for (2) from the error bound condition for the problem (2). As an application, we give a new proof of the convergence of the GPA on the base of obtained results.

1. Auxiliary facts

For a set A define by ∂A and $\operatorname{int} A$ the boundary and the interior of the set A, respectively.

A closed set $A \subset \mathbb{R}^n$ is called *proximally smooth* with constant $R > 0$ [7] if the distance function $\varrho_A(x)$ is continuously differentiable on the set $U_A(R) = \{x \in \mathbb{R}^n \ : 0 < \varrho_A(x) < R\}$. Equivalently, the set $A \subset \mathbb{R}^n$ is proximally smooth with constant $R > 0$ if and only if for any $x \in \partial A$ and $p \in N(A, x) = \{q \in \mathbb{R}^n \mid \exists\, t_0 > 0 : \forall\, t \in (0, t_0)\ \varrho_A(x+tq) = t\|q\|\}$, $\|p\| = 1$, we have $A \cap \operatorname{int} B_R(x + Rp) = \varnothing$ and $P_A x$ is a singleton for all $x \in U_A(R)$.

For example, the Euclidean sphere $\partial B_1(0)$ is proximally smooth with constant 1 and, more generally, the Stiefel manifold $S_{n,k} = \{X \in \mathbb{R}^{n\times k} \mid X^T X = I_k\}$ is also proximally smooth with constant 1 [8, text after Proposition 2] (here n, k are natural with $n \geqslant k$, I_k is the identity $k \times k$ matrix).

In work [9], the next result was proved.

Proposition 1. *Denote by $\Omega \subset S$ the set of stationary points in problem (2) with a real-analytic function and a real analytic set, i.e. for any $x \in \Omega$ we have $-f'(x) \in N(S, x)$. Then there exist $\alpha \in (0, 1]$ and $\mu > 0$ such that for any $x \in S$ the next inequality holds*

$$\varrho(x, \Omega) \leqslant \mu \|P_{T_x} f'(x)\|^{\alpha}. \tag{4}$$

The inequality in Proposition 1 is referred to as the error bound condition. In fact, this is some sort of generalization of the Hoffman lemma in a nonlinear case.

The next proposition gives the rate of decreasing of the function from (2) per step of the gradient projection algorithm.

Proposition 2. *[8, Theorem 2, Corollary 2, Formula (24)] Let S be a C^1 manifold without boundary and a proximally smooth set with constant $\frac{\pi}{2}R$. Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz function with constant L and f' is Lipschitz with constant L_1 in the R-neighborhood $\{x \in \mathbb{R}^n \ :\ \varrho_S(x) < R\}$ of S. Let $t_0 = 1/\left(L_1 + \frac{2L}{R}\right)$, $t \in (0, t_0]$, $x_0 \in S$ and $x_1 = P_S\left(x_0 - tP_{T_{x_0}} f'(x_0)\right)$. Then we have*

$$f(x_0) - f(x_1) \geqslant q(t)\|P_{T_{x_0}} f'(x_0)\|^2, \quad \text{where } q(t) = t - t^2\left(\frac{L_1}{2} + \frac{L}{R}\right). \tag{5}$$

Note that $t_0 < R/L$ and hence $P_S\left(x_0 - tP_{T_{x_0}} f'(x_0)\right)$ is a singleton due to proximal smoothness of the set S.

The next proposition states that any compact smooth manifold without boundary in $\mathbb{R}^n$ is proximally smooth with some constant $R > 0$.

Proposition 3. *[10, Lemma 1.19.3] Let S be a compact smooth manifold without boundary in $\mathbb{R}^n$. Then there exists $R > 0$ such that S is proximally smooth with constant R.*

Unfortunately, it is impossible to estimate R in Proposition 3 in the general case. For some manifolds, we know the precise value of R, for example for most matrix manifolds [11]. Sometimes you can still estimate R, see [12].

We also want to recall that for a Lipschitz differentiable with Lipschitz constant L_1 function f the next quadratic approximations hold [10, Theorem 2.1.2]

$$f(x_0)+(f'(x_0),x-x_0)-\frac{L_1}{2}\|x-x_0\|^2 \leqslant f(x) \leqslant f(x_0)+(f'(x_0),x-x_0)+\frac{L_1}{2}\|x-x_0\|^2,\ \forall\, x,x_0.$$

2. The LPL inequality and convergence of the GPA

Theorem 1. *Assume that conditions of Proposition 1 are fulfilled,* $x_0 \in S$*,* $S_0 = \{x \in S : f(x) \leqslant f(x_0)\}$ *and* $\Omega_0 = \operatorname{Arg}\min\limits_{x\in S} f(x)$*,* $f_* = f(\Omega_0)$*. Suppose that for all* $x \in S_0$ *we have* $\varrho(x,\Omega_0) \leqslant \varrho(x,\Omega\backslash\Omega_0)$*. Then for any* $x \in S_0$ *we have*

$$\|P_{T_x}f'(x)\|^{2\alpha} \geqslant \nu(f(x)-f_*). \tag{6}$$

In (6) $\alpha \in (0,1]$ *is the number from the error bound condition and* $\nu^{-1} = \mu^2\left(\frac{3}{2}L_1 + \frac{L_0}{R}\right)$*. Here* $\mu > 0$ *is from the error bound condition,* L_1 *is a Lipschitz constant of* f' *on* $U_S(R)$*,* $L_0 = \max\limits_{x\in\Omega_0}\|f'(x)\|$ *and* R *is constant of proximal smoothness for the set* S*.*

Proof by Proposition 1 Formula (4) holds. By Proposition 3 the set S is proximally smooth with some constant $R > 0$.

Fix $x \in S_0$ and $x_* \in P_\Omega x = P_{\Omega_0}x$.

If $f'(x_*) = 0$, then $(f'(x_*), x - x_*) = 0$. Further assume that $f'(x_*) \neq 0$.

Consider case 1, when $x \in S_0 \cap U_{\Omega_0}(R)$. From the supporting principle for proximally smooth sets and necessary condition of extremum $\pm f'(x_*) \in N(S,x_*)$ we have $\operatorname{int} B_R\left(x_* \pm R\frac{f'(x_*)}{\|f'(x_*)\|}\right)\cap S = \varnothing$. Define $H = \{x \in \mathbb{R}^n : (f'(x_*), x-x_*) = 0\}$ and $z = P_H x$. Then from the inclusion $x \in U_{\Omega_0}(R)$ we obtain that $\|z - x_*\| \leqslant \|x - x_*\| = \varrho_{\Omega_0}(x) < R$ and

$$\|x-z\| \leqslant R - \sqrt{R^2 - \|z-x_*\|^2} \leqslant \frac{\|z-x_*\|^2}{R} \leqslant \frac{\|x-x_*\|^2}{R},$$

$$(f'(x_*), x-x_*) = \|f'(x_*)\|\cdot\|x-z\| \leqslant \frac{L_0}{R}\|x-x_*\|^2. \tag{7}$$

Consider case 2, when $x \in S_0\backslash U_{\Omega_0}(R)$. Then $\|x - x_*\| = \varrho_{\Omega_0}(x) \geqslant R$ and hence

$$(f'(x_*), x-x_*) \leqslant \|f'(x_*)\|\cdot\|x-x_*\| \leqslant \frac{L_0}{R}\|x-x_*\|^2.$$

Thus for any $x \in S_0$ we have condition (7).

From the quadratic approximation of the Lipschitz differentiable function f we get

$$f(x) - f(x_*) \leqslant (f'(x), x-x_*) + \frac{L_1}{2}\|x-x_*\|^2,$$

$$(f'(x), x-x_*) = (f'(x) - f'(x_*), x-x_*) + (f'(x_*), x-x_*) \leqslant$$

$$\leqslant L_1\|x-x_*\|^2 + \frac{L_0}{R}\|x-x_*\|^2 = C\|x-x_*\|^2, \quad C = L_1 + \frac{L_0}{R}.$$

Finally we have, taking in mind $\|x - x_*\| = \varrho_\Omega(x)$,

$$f(x) - f(x_*) \leqslant \left(C + \frac{L_1}{2}\right)\|x-x_*\|^2 \leqslant \mu^2 D\|P_{T_x}f'(x)\|^{2\alpha}, \quad D = C + \frac{L_1}{2}. \qquad \square$$

On the base of the LPL inequality we prove the convergence of the GPA.

Theorem 2. *Assume that conditions of Proposition 2 and Theorem 1 are fulfilled. Then the iterations (3) give the sequence $\{x_k\}$ with $f(x_k) \to f(\Omega)$. The rate of convergence is given by Formula (8).*

Proof. Fix $k \geqslant 1$. From Formula (5) we have

$$f(x_k) - f(x_{k+1}) \geqslant q(t)\|P_{T_{x_k}} f'(x_k)\|^2$$

and $f(x_k) \leqslant f(x_0)$, i.e. $x_k \in S_0$. By Theorem 1 there exist numbers $\nu > 0$ and $\beta \in (0, 2]$ with

$$\|P_{T_x} f'(x)\|^\beta \geqslant \nu(f(x) - f_*) \quad \forall\, x \in S_0.$$

Combining the two last formulae we get

$$f(x_k) - f(x_{k+1}) \geqslant q(t)\nu^{\frac{2}{\beta}}(f(x) - f_*)^{\frac{2}{\beta}}.$$

Put $\varphi_k = f(x_k) - f_*$. Then

$$\varphi_k - \varphi_{k+1} \geqslant q(t)\nu^{\frac{2}{\beta}}\varphi_k^{\frac{2}{\beta}}, \quad \varphi_{k+1} \leqslant \varphi_k - q(t)\nu^{\frac{2}{\beta}}\varphi_k^{\frac{2}{\beta}}. \tag{8}$$

□

Formula (8) ensures the convergence. For example, for $\beta = 2$, we have the linear rate of convergence. Moreover, if $\beta = 2$ then $1 - \nu q(t) \geqslant 0$. If $a = \frac{2}{\beta} - 1 > 0$ then by [13, Chapter 2, Lemma 6]

$$\varphi_k \leqslant \frac{\varphi_0}{\left(1 + aq(t)\nu^{\frac{2}{\beta}}\varphi_0^a k\right)^{1/a}}.$$

Notice that Theorems 1 and 2 become constructive if we have information about the error bound condition (4) and the constant of proximal smoothness for S. For example, if f and S are an arbitrary C^2 function and a compact manifold, respectively, and Ω is a finite set then, under some assumption of nondegeneracy, Formula (4) holds with $\alpha = 1$ [14, Theorem 1].

If the value R is unknown then there can be a problem with choosing the step size t. In this case, the GPA with Armijo's step size can be considered [11].

References

1. Lee J. M. Manifolds and differential geometry. *Graduate Studies in Mathematics*. Rhode Island, AMS Providence, 2009. Vol. 107. 671 p. https://doi.org/10.1090/gsm/107
2. Lojasiewicz S. Sur le probléme de la division. *Studia Mathematica*, 1959, vol. 18, pp. 87–136. https://doi.org/10.4064/sm-18-1-87-136
3. Balashov M. V., Polyak B. T., Tremba A. A. Gradient projection and conditional gradient methods for constrained nonconvex minimization. *Numerical Functional Analysis and Optimization*, 2020, vol. 41, iss. 7, pp. 822–849. https://doi.org/10.1080/01630563.2019.1704780
4. Karimi H., Nutini J., Schmidt M. Linear convergence of gradient and proximal-gradient methods under the Polyak – Lojasiewicz condition. In: Frasconi P., Landwehr N., Manco G., Vreeken J. (eds.) *Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2016*. Lecture Notes in Computer Science, vol. 9851. Cham, Springer, 2016, pp. 795–811. https://doi.org/10.1007/978-3-319-46128-1_50

5. Schneider R., Uschmajew A. Convergence results for projected line-search methods on varieties of low-rank matricies via Lojasiewicz inequality. *SIAM Journal on Optimization*, 2015, vol. 25, iss. 1, pp. 622–646. https://doi.org/10.1137/140957822
6. Merlet B., Nguyen T. N. Convergence to equilibrium for discretizations of gradient-like flows on Riemannian manifolds. *Differential Integral Equations*, 2013, vol. 26, iss. 5/6, pp. 571–602. https://doi.org/10.57262/die/1363266079
7. Vial J.-Ph. Strong and weak convexity of sets and functions. *Mathematics of Operations Research*, 1983, vol. 8, iss. 2, pp. 231–259. https://doi.org/10.1287/moor.8.2.231
8. Balashov M. V. The gradient projection algorithm for smooth sets and functions in nonconvex case. *Set-Valued and Variational Analysis*, 2021, vol. 29, pp. 341–360. https://doi.org/10.1007/s11228-020-00550-4
9. Balashov M. V. Stability of minimization problems and the error bound condition. *Set-Valued and Variational Analysis*, 2022, vol. 30, pp. 1061–1076. https://doi.org/10.1007/s11228-022-00634-3
10. Ivanov G. E. *Slabo vypuklye mnozhestva i funktsii: teoriya i prilozhenie* [Weakly Convex Sets and Functions: Theory and Application]. Moscow, Fizmatlit, 2006. 351 p. (in Russian).
11. Balashov M. V., Kamalov R. A. The gradient projection method with Armijo's step size on manifolds. *Computational Mathematics and Mathematical Physics*, 2021, vol. 61, iss. 1, pp. 1776–1786. https://doi.org/10.1134/S0965542521110038
12. Adly S., Nacry F., Thibault L. Preservation of prox-regularity of sets with applications to constrained optimization. *SIAM Journal on Optimization*, 2016, vol. 26, iss. 1, pp. 448–473. https://doi.org/10.1137/15M1032739
13. Polyak B. T. *Introduction to Optimization*. New York, Optimization Software, 1987. 464 p.
14. Balashov M. V., Tremba A. A. Error bound conditions and convergence of optimization methods on smooth and proximally smooth manifolds. *Optimization: A Journal of Mathematical Programming and Operations Research*, 2020, vol. 71, iss. 3, pp. 711–735. https://doi.org/10.1080/02331934.2020.1812066

Поступила в редакцию / Received 20.08.2022
Принята к публикации / Accepted 27.10.2022
Опубликована / Published 01.03.2023