УДК 519.85, 517.97

УСЛОВИЕ ПОЛЯКА-ЛОЯСЕВИЧА НА ГЛАДКОЙ ПОВЕРХНОСТИ

М. В. Балашов

Институт проблем управления им. В.А. Трапезникова РАН

Россия, 117997, Москва, Профсоюзная ул., 65

E-mail: balashov73@mail.ru

Ключевые слова: условие Поляка-Лоясевича, прокс-регулярность, метод проекции градиента.

Аннотация: Сформулирован аналог условия Поляка-Лоясевича на гладкой поверхности. Показано, что это условие обеспечивает сходимость метода проекции градиента для минимизации функции с непрерывным по Липшицу градиентом на гладкой прокс-регулярной поверхности с линейной скоростью. Функция не предполагается выпуклой. Рассмотрены примеры.

1. Введение

Метод проекции градиента (МПГ) появился в работах [1] (без оценки скорости сходимости) и [2] (с оценкой скорости сходимости). МПГ показал себя эффективным для решения разных экстремальных задач и связанных с ними вопросов [2, 3]. Выпуклость (множества и функции) и гладкость (функции) являются ключевыми свойствами для эффективной работы метода. Например, если функция сильно выпуклая с непрерывным по Липшицу градиентом, а множество выпуклое, то МПГ сходится с линейной скоростью. В некоторых случаях выпуклость функции может быть заменена на т.н. свойство инвексивности ("invexity") [4]. Последнее в частности означает, что любая стационарная точка функции является точкой глобального экстремума. Мы также хотим заметить, что можно требовать дополнительные свойства выпуклости для множества в МПГ (например, сильную выпуклость множества) и рассматривать при этом просто выпуклые или даже невыпуклые гладкие функции [5, 6].

Одно из первых условий инвексивности — условие Поляка-Лоясиевича (ПЛ) [7, 8]. Будем говорить, что условие ПЛ имеет место для ограниченной снизу функции $f : \mathbb{R}^d \supset U \to \mathbb{R}$, если существует число $\mu > 0$ такое, что $\|f'(x)\|^2 \geqslant \mu(f(x) - f_0)$ для всех $x \in U$. Здесь $f_0 = \inf\limits_{y\in U} f(y)$. Важным примером невыпуклой функции, удовлетворяющей условию ПЛ, является функция $f(x) = \|g(x)\|^2$, где $g : \mathbb{R}^d \to \mathbb{R}^m$ ($m < d$) непрерывно дифференцируемое отображение с матрицей Якоби полного ранга.

Указанное условие ПЛ позволило доказать линейную скорость сходимости метода градиентного спуска в задаче безусловной минимизации функции с липшицевым градиентом [7].

Мы планируем рассмотреть условие ПЛ на гладкой прокс-регулярной m-мерной поверхности S в $\mathbb{R}^d$. Гладкость понимается как C^1. В частности, в каждой точке поверхности есть касательная плоскость. Замкнутое множество $S \subset \mathbb{R}^d$ называется *прокс-регулярным* (или *проксимально гладким, слабо выпуклым*) с константой R, если функция расстояния $\varrho_S(x)$ от точки $x \in \mathbb{R}^d$ до множества S непрерывно дифференцируема на множестве

$$U_S(R) = \{x \in \mathbb{R}^d \mid 0 < \varrho_S(x) < R\}.$$

Свойства таких множеств описаны в [9–11].

Эквивалентным определению прокс-регулярности множества S является опорное условие: для всякой граничной точки $x \in \partial S$ и всякой единичной нормали $n(x) \in N(S,x)$ (здесь $N(S,x)$ — любой нормальный конус к S в точке $x \in \partial S$) выполняется равенство

$$A \cap \operatorname{int} B_R(x + Rn(x)) = \varnothing.$$

Здесь $B_R(x)$ замкнутый шар радиуса R с центром в точке x.

Заметим, что $d-1$-мерная поверхность является прокс-регулярной с константой R тогда и только тогда, когда для непрерывной единичной нормали $S \ni x \to n(x)$ выполнено неравенство $\|n(x_1) - n(x_2)\| \leqslant R^{-1}\|x_1 - x_2\|$. Для поверхностей размерности $m < d-1$ часто возможно оценить константу проксимальной регулярности.

2. Условие Поляка-Лоясевича и метод проекции градиента

Определим $\mathcal{L}_f(\alpha) = \{x \in \mathbb{R}^d \mid f(x) \leqslant \alpha\}$,

Определение 1. *Пусть S C^1 - гладкая прокс-регулярная поверхность. Будем говорить, что условие ПЛ выполняется на поверхности S, если $f_0 = \inf_S f > -\infty$ и существует $\mu > 0$, что для всякой точки $x \in S \cap \mathcal{L}_f(\alpha)$*

$$\|P_x f'(x)\|^2 \geqslant \mu\left(f(x) - f_0\right).$$

При этом мы требуем, чтобы край поверхности S не содержался в $\mathcal{L}_f(\alpha)$. Эту ситуацию мы будем упрощенно называть "поверхностью без края". Здесь P_x метрический проектор на касательное подпространство T_x к поверхности S в точке x.

Условие ПЛ позволяет доказать сходимость метода проекции градиента для случая прокс-регулярной поверхности и функции с липшицевым градиентом.

Теорема 1. *Пусть S гладкая и проксимально регулярная поверхность с константой R без края, $x_0 \in S$, $t > 0$. Предположим, что функция $f : \mathbb{R}^d \to \mathbb{R}$ имеет следующие свойства*
1) f' непрерывна по Липшицу с константой L_1,
2) $L = \max\limits_{x \in S} \|f'(x)\|$,
Предположим, что $t\|f'(x_0)\| \leqslant \frac{\sqrt{3}}{2}R$. Пусть $z = P_{x_0+T_{x_0}}(x_0 - tf'(x_0))$, $x_1 = S \cap (z + T_{x_0}^{\perp}) \cap B_R(x_0)$. Тогда

$$f(x_1) - f(x_0) \leqslant -\|P_{x_0} f'(x_0)\|^2 \left[t - t^2\left(\frac{L}{R} + \frac{L_1}{2} + t^2\frac{L_1 L^2}{2R^2}\right)\right]. \tag{1}$$

В теореме $P_{x_0+T_{x_0}}$ — метрический проектор на касательное подпространство $x_0 + T_{x_0}$. Заметим также, что точка x_1 в теореме *однозначно* определяется.

Отметим, что в теореме описан один шаг метода проекции градиента. Если на каждом шаге выполнены условия теоремы, то МПГ сходится со скоростью геометрической прогрессии. Доказательство этого факта повторяет рассуждения из [7].

3. Проверка условия Поляка-Лоясевича

Рассмотрим, как проверять условие ПЛ. Оказывается, очень часто задачу можно свести к проверке "безусловного"условия ПЛ.

Лемма 1. *Пусть $D \subset \mathbb{R}^m$ есть замыкание измеримой области, $\varphi : D \to \mathbb{R}^d$ ($m < d$), $\varphi \in C^1(D)$, $\operatorname{rg} \varphi'(u) = m$ для всех $u \in D$. Предположим, что $\varphi : D \to S$ есть биекция и $S = \varphi(D)$ является гладкой поверхностью.*

Допустим, что функция f удовлетворяет условию ПЛ на поверхности S с константой μ. Тогда функция $h(u) = f(\varphi(u))$ удовлетворяет условию ПЛ на множестве D с константой $\alpha_0^2\mu$, где $\alpha_0 > 0$ такое число, что $\|\varphi'(u)\ell\| \geqslant \alpha_0\|\ell\|$ для всех $\ell \in \mathbb{R}^m$ и $u \in D$.

Доказательство. Из компактности множества D и условия $\operatorname{rg} \varphi'(u) = m$ для всех $u \in D$ легко следует, что $\alpha_0 > 0$.

Зафиксируем $x \in S$ и соответствующую точку $u \in D$ со свойством $x = \varphi(u)$. Имеем

$$h'(u) = f'_x(\varphi(u))\varphi'(u), \tag{2}$$

где $f'_x(\varphi(u)) = f'(x)|_{x=\varphi(u)}$.

Выберем такой вектор $\ell \in \mathbb{R}^m$, что $P_x f'(x) = \varphi'(u)\ell$ (ℓ зависит от x). Тогда $(h'(u), \ell) = (f'_x(\varphi(u)), \varphi'(u)\ell)$ и

$$\|h'(u)\|\|\ell\| \geqslant |(h'(u), \ell)| = |(f'_x(\varphi(u)), P_{T_x} f'_x(\varphi(u)))| = \|P_{T_x} f'_x(\varphi(u)))\|^2.$$

В силу определения α_0 мы получаем

$$\|P_x f'_x(\varphi(u)))\| = \|\varphi'(u)\ell\| \geqslant \alpha_0\|\ell\|.$$

Отсюда $\|\ell\| \leqslant \frac{1}{\alpha_0}\|P_x f'_x(\varphi(u)))\|$ и

$$\|h'(u)\| \geqslant \alpha_0\|P_x f'_x(\varphi(u)))\|.$$

□

Лемма 2. *Пусть $D \subset \mathbb{R}^m$ есть замыкание измеримой области , $\varphi : D \to \mathbb{R}^d$ ($m < d$), $\varphi \in C^1(D)$. Предположим, что $\varphi : D \to S$ есть биекция и $S = \varphi(D)$ является гладкой поверхностью.*

Пусть ψ есть гладкое обратное отображение для $\varphi : D \to S$: $\psi(\varphi(u)) = u$ при всех $u \in D$. Допустим, что функция h удовлетворяет условию ПЛ на множестве D с константой μ. Тогда функция $f(x) = h(\psi(x))$ удовлетворяет условию ПЛ на поверхности S с константой μ/β_0^2, где $\beta_0 = \sup\limits_{u \in D} \|\varphi'(u)\|$.

Доказательство. Выберем $u \in D$ и $x = \varphi(u)$. По формуле (2) получаем

$$h'(u) = h'_u(\psi(x)) = f'_x(x)\varphi'(u).$$

Пусть $\ell \in \mathbb{R}^m$ есть такой вектор, что $\|\ell\| = 1$ и $(h'(u), \ell) = \|h'(u)\|$ (ℓ зависит от u). Тогда

$$\|h'(u)\| = (h'(u), \ell) = (f'_x(x), \varphi'(u)\ell) = \|f'_x(x)\|\|\varphi'(u)\ell\| \cos\theta \leqslant \beta_0\|f'_x(x)\| \cos\theta,$$

где θ есть угол между векторами $f'_x(x)$ и $\varphi'(u)\ell \in T_x$. Определим через γ угол между $f'_x(x)$ и T_x, в силу определения $\gamma \leqslant \theta$ и

$$\|h'(u)\| \leqslant \beta_0\|f'_x(x)\| \cos\gamma = \beta_0\|P_{T_x} f'_x(x)\|,$$

то есть $\frac{1}{\beta_0}\|h'(u)\| \leqslant \|P_{T_x} f'_x(x)\|$. □

Рассмотрим пример. Пусть $f(x) = (x, Ax)$, где A — симметричная матрица с собственными значениями $\lambda_1 < \lambda_2 \leqslant \cdots \leqslant \lambda_d$. Определим константу μ в условии ПЛ функции f на множестве $S_\varepsilon = \{\|x\| = 1 \mid x_1 \geqslant \varepsilon\}$. Мы рассматриваем координаты в базисе из собственных векторов матрицы A.

Решим более общую задачу. Пусть функция f задана на множестве S_ε. Допустим, что функция

$$h(x_2, \dots, x_d) = f\left(\sqrt{1 - \sum_{k=2}^{k=d} x_k^2}, x_2, \dots, x_d\right)$$

удовлетворяет условию ПЛ на множестве $\sum\limits_{k=2}^{k=d} x_k^2 \leqslant 1 - \varepsilon^2$ (в $\mathbb{R}^{d-1}$) с константой μ_0. Рассмотрим (для $u = (x_2, \dots, x_d)$) вектор-функцию

$$\varphi(u) = \left(\sqrt{1 - \sum_{k=2}^{d} x_k^2}, x_2, \dots, x_d\right)$$

и матрицу Якоби (d строк, $d-1$ столбцов)

$$\varphi' = \begin{pmatrix} -\frac{x_2}{\sqrt{1-\sum\limits_{k=2}^{d} x_k^2}} & \dots & -\frac{x_{d-1}}{\sqrt{1-\sum\limits_{k=2}^{d} x_k^2}} & -\frac{x_d}{\sqrt{1-\sum\limits_{k=2}^{d} x_k^2}} \\ 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & 1 & 0 \\ 0 & \dots & 0 & 1 \end{pmatrix}$$

Пусть $\|\ell\| = 1$, тогда

$$\|\varphi'\ell\|^2 = \left(\sum_{k=2}^{d} \frac{x_k\ell_k}{\sqrt{1 - \sum\limits_{j=2}^{d} x_j^2}}\right)^2 + \sum_{k=2}^{d} \ell_k^2 \leqslant \sum_{k=2}^{d}\left(\frac{x_k}{\sqrt{1 - \sum\limits_{j=2}^{d} x_j^2}}\right)^2 + 1 \leqslant \frac{1-\varepsilon^2}{\varepsilon^2} + 1 = \frac{1}{\varepsilon^2},$$

значит в лемме 2 $\beta_0 = \frac{1}{\varepsilon}$. Поэтому по лемме 2 константа $\mu = \mu_0\varepsilon^2$.

Для случая квадратичной функции $f(x) = (x, Ax)$ с помощью замены $x_1^2 = 1 - \sum\limits_{k=2}^{d} x_k^2$ легко видеть, что $h(x) = \sum\limits_{k=2}^{d}(\lambda_k - \lambda_1)x_k^2$. Для h как функции в $\mathbb{R}^{d-1}$ легко вычисляется, что $\mu_0 = 4(\lambda_2 - \lambda_1)$. Для f по лемме 2 $\mu = 4(\lambda_2 - \lambda_1)\varepsilon^2$. Последний результат получен А. Тремба другим способом.

Работа выполнена при поддержке Российского фонда фундаментальных исследований (18-01-00209).

Список литературы

1. Goldstein A.A. Convex programming in Hilbert space // Bull. Amer. Math. Soc. 1964. Vol. 70, No. 5. P. 709-710.
2. Levitin E.S., Polyak B.T. Constrained minimization methods // Zh. Vychisl. Mat. Mat. Fiz. 1966. Vol. 6, No. 5. P. 787-823.
3. Polyak B.T. Introduction to optimization. M.: Science, 1983.
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. Springer.
5. Balashov M.V., Golubev M.O. About the Lipschitz property of the metric projection in the Hilbert space // Journal of Mathematical Analysis and Applications. 2012. Vol. 394, No. 2. P. 545-551.
6. Balashov M.V. Maximization of a function with Lipschitz continuous gradient // Journal of Mathematical Sciences. 2015. Vol. 209, No. 1. P. 12-18.
7. Polyak B.T. Gradient methods for minimizing functionals // Zh. Vychisl. Mat. Mat. Fiz. 1963. Vol. 3, No. 4. P. 643-653.
8. Lojasiewicz S. A Topological Property of Real Analytic Subsets // Coll. du CNRS. Les équations aux dérivées partielles. 1963. Vol. 117. P. 87-89. (in French)
9. Vial J.-Ph. Strong and weak convexity of sets and functions, Mathematics of Operations Research. 1983. Vol. 8, No. 2. P. 231-259.
10. Clarke F.H., Stern R.J., Wolenski P.R. Proximal smoothness and lower–C^2 property // J. Convex Anal. 1995. Vol. 2, No. 1-2. P 117-144.
11. Balashov M.V., Ivanov G.E. Weakly convex and proximally smooth sets in Banach spaces // Izv. RAN. Ser. Mat. 2009. Vol. 73, No. 3. P. 23-66.