مجلة جامعة طرطوس للبحوث والدراسات العلمية_ سلسلة العلوم الأساسية المجلد (6) العدد (2) 2022
Tartous University Journal for Research and Scientific Studies –Basic Sciences Series Vol. (6) No. (2) 2022

دراسة التحدب الضعيف في مسائل الأمثليات

لما رزوق*

(تاريخ الإيداع 6/ 3/ 2022 – تاريخ النشر 27 / 4/ 2022)

□ ملخّص □

مع التقدم الأخير في خوارزميات علوم الحاسوب و الأمثليات (التحسين)، يدرس الباحثون في الغالب مسائل الأمثليات المحدبة نظراً لما تملكه من بنية جيدة وتسمح بتقدير معدل التقارب للخوارزميات المقترحة.

عند إنشاء خوارزميات فعالة لمسائل الأمثليات، هناك عاملان يعتبران مهمين للغاية: التحدب ومستوى النعومة (الملاسة).

في هذا البحث قدمنا مبرهنة أثبتنا فيها أن الدالة تكون محدبة بضعف إذا حققت هذه الدالة شرط هولدر للتدرج المستمر من أجل أي قيمة لوسيط التنعيم $\nu \in (0,1]$. وهذا يساعد في دراسة مسائل أعم من الموجودة في أعمال سابقة [7;13;16].

الكلمات المفتاحية: الدالة الناعمة(الملساء)، الدالة ذات التحدب الضعيف، مسألة النقطة السرجية، شرط هولدر.

*معيدة حاصلة على شهادة الماجستير من MIPT – قسم الرياضيات – كلية العلوم – جامعة طرطوس – طرطوس – سوريا.

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الأساسية المجلد (6) العدد (2) 2022
Tartous University Journal for Research and Scientific Studies –Basic Sciences Series Vol. (6) No. (2) 2022

Study of the weak-convexity in optimization problems

Lama Y. Razouk *

(Received 6/3/2022.Accepted 27/4/2022)

□ABSTRACT □

With recent advancements in computing and optimization algorithms, researchers mostly considered convex optimization problems since they have good structure and allow to estimate rate of convergence for proposed algorithms.

In the construction of efficient schemes for optimization problems, there are two factors considered very important: Convexity and level of Smoothness.

In this research, we prove a theorem which can give a weakly-convex function when it has

ν-Hölder-continuous gradient. This work studies more general than in the existing literature problems [7;13;16].

Key Words: Smooth function, Weakly-Convex function, Saddle point problem, Hölder condition.

*Assistant with a master degree from MIPT− Department of Mathematics− Faculty of Science− Tartous University− Tartous− Syria.

1 Introduction:

Convex optimization has applications in different fields, such as automatic control systems, communications and signal processing [10], electronic circuit design [4], data analysis and modeling, finance and structural optimization, where the approximation concept has proven to be efficient [4; 15].

Recently, a wide study has been done to find properties for the function in saddle point (minimax) problems of the form:

$$\min_{x\in X}\max_{y\in Y} f(x,y)\ ;\ f: X\times Y\to\mathbb{R} \qquad (1)$$

where f is smooth function, $X\subseteq\mathbb{R}^n$ and $Y\subseteq\mathbb{R}^m$, $n,m\in\mathbb{N}$.

This problem has applications in various domains such as machine learning [6; 11], optimization [2], statistics [1], mathematics [8], and game theory [12]. Given the importance of these problems, there are several studies about different algorithms and their convergence properties.

While most theoretical studies focus on the smooth function, several real-world problems fall outside this class. In our research, we prove a "general theorem", which gives the weak-convexity of $\max_{y\in Y} f(x,y)$ in all level smoothness of f.

Basic related works: In [16], they studied the smooth nonconvex concave minimax problems

and they measured the convergence to an approximate (first order stationary points) FOSP [13;7] of this problem which requires weak-convexity of $\max_{y\in Y} f(x,y)$. So, they had proved *lemma* (1) (below) to guarantee weak convexity of $\max_{y\in Y} f(x,y)$ given smoothness of f.

2 The Importance and the Aim of the research:

The importance of this research appears with the need to weak-convexity of the used function in minimax problems where the function has different levels of smoothness depend on the parameter $\nu\in(0,1]$.

The aim of this research is to prove that f is weakly-convex when it has ν-Hölder-continuous gradients.

3 Research Method and its Resources:

We use known mathematical definitions and notions.

$\mathbb{R}^q$ is the real vector space of dimension q, where $q\in\mathbb{N}$.

4 Preliminaries and Notations for the minimax problems:

Definition 1: [16] A function $f(x,y)$ is said to be L-*smooth*, for some $L > 0$, if:

$$\max\{\| \nabla_x f(x,y) - \nabla_x f(x',y') \|, \| \nabla_y f(x,y) - \nabla_y f(x',y') \|\} \leq L(\| x - x' \| + \| y - y' \|), \forall x, x' \in X, y, y' \in Y \quad (2)$$

Definition 2: [5] Let the function f be a convex and subdifferentiable on X, then for the constant $L_v < +\infty$, $\forall\, x, x' \in X$, its gradient with respect to x satisfies the *Hölder Continues Condition* if:

$$\| \nabla_x f(x',y) - \nabla_x f(x,y) \|_* \leq L_v \| x - x' \|^v,$$

(3)

where $v \in (0,1]$.

We note that when $v = 1$, f will be a smooth function and when $v < 1$, we get a lower level of smoothness.

Let us denote the function $g(\cdot)$ and the mapping $y^*(\cdot)$, that given in the set X, as follows

$$g(x) = \max_{y \in Y} f(x,y), y^*(x) = \underset{y \in Y}{\operatorname{argmax}} f(x,y), \forall x \in X. \quad (4)$$

Since $g(x)$ could be non-smooth, $\nabla g(x)$ might not even be defined, we need first to generalize the notion of gradient[1] for non-smooth function:

Definition 3: [16] The *Fréchet sub-differential* of a function $g(\cdot)$ at x is defined as the set,

$$\partial g(x) = \{u | \liminf_{x' \to x} \frac{g(x') - g(x) - \langle u, x' - x \rangle}{\|x - x'\|} \geq 0\} \quad . \quad (5)$$

Definition 4: [16] A function $g : X \to \mathbb{R} \cup \{\infty\}$ is L-*weakly convex* if

$$g(x') \geq g(x) - \langle u_{x'}, x' - x \rangle - \frac{L}{2} \qquad \|x' - x\|^2,$$

$\| x' - x \|^2$,   (6)

for all Fréchet subgradients $u_x \in \partial g(x), \forall\, x, x' \in X$.

5 The Discussion of our main theorem:

The following lemma guarantees weak-convexity of g given smoothness of f where $X = \mathbb{R}^n$ and $Y \subset \mathbb{R}^m$ is convex compact set:

[1]For non-smooth problems, i.e., when the objective function f is non-differentiable, we cannot use the gradient. For the non-differentiable functions, there is an important notion, which is the generalization of the gradients for differentiable functions. This notion is *subgradients*.

Lemma (1): [16] Let $f(.,y)$ be continuous and Y be compact. Then $g(x)=\max_{y\in Y} f(x,y)$ is L −weakly convex, if f is L−weakly convex in x (Definition 1), or if f is L−smooth in x.

We generalize this fact making lemma (1) a special case when $\nu=1$. Since in the following theorem (1) the gradients of f are ν−Hölder−continuous for $\nu\in(0,1]$ and that introduce the universal case of smoothness for the used function f.

The Main General Theorem (1): Let $f(.,y)$ be continuous, Y be compact set. If f has ν−Hölder−continuous gradient for some $\nu\in(0,1]$, or if f is L−weakly convex in x w.r.t. norm $\|\cdot\|$, then g is L−weakly convex in x w.r.t. norm $\|\cdot\|$. Also, we have

$$\partial g(x)=\operatorname{conv}\left\{\partial_x f(x,y^*(x))\,\middle|\, y^*(x)=arg\max_{y\in Y} f(x,y)\right\},\forall x\in X. \qquad (7)$$

where $\operatorname{conv}\{\cdot\}$ means the convex hull of a given set [14].

Proof:

Since f has ν−Hölder−continuous gradient with respect to x, then we have

$$\|\nabla_x f(x',y)-\nabla_x f(x,y)\|_*\le L_\nu\|x-x'\|^\nu,\quad \forall x,x'\in X. \qquad (8)$$

By notice that $f(x',y)-f(x,y)=\int_0^1\langle\nabla f(x+t(x'-x),y),x'-x\rangle\,dt$,

thus

$$f(x',y)-f(x,y)-\langle\nabla f(x,y),x'-x\rangle$$

$$\le\int_0^1\langle\nabla f(x+t(x'-x),y)-\nabla f(x,y),x'-x\rangle\,dt,$$

therefore

$$|f(x',y)-f(x,y)-\langle\nabla f(x,y),x'-x\rangle|$$
$$\le\int_0^1|\langle\nabla f(x+t(x'-x),y)-\nabla f(x,y),x'-x\rangle|\,dt$$
$$\le\int_0^1\|\nabla f(x+t(x'-x),y)-\nabla f(x,y)\|\|x'-x\|\,dt$$
$$\le\int_0^1\|\nabla f(x+t(x'-x),y)-\nabla f(x,y)\|\|x'-x\|\,dt$$
$$\le^{(8)}\int_0^1 L_\nu\|t(x'-x)\|^\nu\|x'-x\|\,dt$$
$$=L_\nu\|x'-x\|^{1+\nu}\int_0^1 t^\nu dt=\frac{L_\nu}{1+\nu}\|x'-x\|^{1+\nu}.$$

Thus, we find that

$$f(x',y) \geq f(x,y) + \langle \nabla f(x,y),\ x' - x \rangle - \frac{L_\nu}{1+\nu} \| x' - x \|^{1+\nu}.$$

{\displaystyle X}

Here we have two cases:

- In case $\| x' - x \| \geq 1$:

$$1 + \nu = \frac{(1+\nu)^2}{1+\nu} = \frac{1+2\nu}{1+\nu} + \frac{\nu^2}{1+\nu} < \frac{2(1+\nu)}{1+\nu} + \frac{\nu^2}{1+\nu}.$$

Therefore

$$f(x',y) \geq f(x,y) + \langle \nabla f(x,y),\ x' - x \rangle - \frac{L_\nu}{1+\nu} \| x' - x \|^{2+\frac{\nu^2}{1+\nu}}$$

$$= f(x,y) + \langle \nabla f(x,y),\ x' - x \rangle - \frac{L_\nu}{1+\nu} \| x' - x \|^{\frac{\nu^2}{1+\nu}} \| x' - x \|^2,$$

where $L = 2\frac{L_\nu}{1+\nu} \| x' - x \|^{\frac{\nu^2}{1+\nu}} \leq 2\frac{L_\nu}{1+\nu} D_0^{\frac{\nu^2}{1+\nu}}$; $D_0 = sup_{x,x' \in X} \| x' - x \|,\ \nu \in (0,1]$.

Then

$$f(x',y) \geq f(x,y) + \langle \nabla f(x,y),\ x' - x \rangle - \frac{L_\nu}{1+\nu} D_0^{\frac{\nu^2}{1+\nu}} \| x' - x \|^2,$$

it means that f is an $L \leq 2\frac{L_\nu}{1+\nu} D_0^{\frac{\nu^2}{1+\nu}}$ – weakly convex.

- In case $\| x' - x \| < 1$:

$$1 + \nu = 2 - 1 + \nu > 2 - 1 - \nu = 2 - (1 + \nu),$$

therefore, we have

$$\| x' - x \|^{1+\nu} < \| x' - x \|^{2-(1+\nu)}.$$

So
$$f(x',y) \geq f(x,y) + \langle \nabla f(x,y),\ x' - x \rangle - \frac{L_\nu}{1+\nu} \| x' - x \|^{2-(1+\nu)}$$

$$= f(x,y) + \langle \nabla f(x,y),\ x' - x \rangle - \frac{L_\nu}{1+\nu} \| x' - x \|^{-(1+\nu)} \| x' - x \|^2,$$

where here we consider

$L = 2 \frac{L_\nu}{1+\nu} \| x' - x \|^{-(1+\nu)} \leq 2 \frac{L_\nu}{1+\nu} D_0^{-(1+\nu)}$; $D_0 = sup_{x,x' \in X} \| x' - x \|,\ \nu \in (0,1]$.

Thus, we only need to prove the case of L–weakly convex $f(.,y)$.

Since $f(.,y)$ is L–weakly convex and $u_{x,y} \in \partial_x f(x,y)$, we get that,

$$f(x',y) \geq f(x,y) + \langle u_{x,y},\ x' - x \rangle - \frac{L}{2} \| x' - x \|^2$$

Therefore

$f\ (x',y) + \frac{L}{2}\| x' \|^2 \geq\ f\ (x,y) + \frac{L}{2}$ This means that $\tilde{f}(x,t) := f\ (x,y) +\ \frac{L}{2}\ \| x \|^2$ is convex, since $\partial_x \tilde{f}(x,t) := \partial_x f\ (x,y) + Lx$ [9].

Let $\tilde{g}(x) = \max_{y \in Y} \tilde{f}\ (x,y)$. Since $\tilde{f}\ (x,y)$ is convex in x and smooth, and Y is compact set, we use Danskin's theorem [3] to prove to that

$$\partial \tilde{g}\ (x)\ =\ \text{conv}\left\{\partial_x \tilde{f}\big(x,y^*(x)\big)\ \middle|\ y^*(x) = arg \max_{y \in Y} \tilde{f}(x,y)\right\}, \forall\ x\ \in\ X$$

$$\Rightarrow\ \partial\ g\ (x) +\ \text{L}x\ =\ \text{conv}\left\{\partial_x f\big(x,y^*(x)\big) + Lx\ \middle|\ y^*(x) = arg \max_{y \in Y} f\,(x,y)\right\},$$

$$\Rightarrow\ \partial g(x)\ =\ \text{conv}\left\{\partial_x f\big(x,y^*(x)\big)\middle|\ y^*(x) = arg \max_{y \in Y} f(x,y)\right\}, \tag{9}$$

where the last two steps are true because of the facts

$$\partial \tilde{g}(x) = \partial g(x) + Lx\ ,\ \partial_x \tilde{f}(x,y) = \partial_x f(x,y) + Lx$$

and $$arg \max_{y \in Y} \tilde{f}(x,y) = arg \max_{y \in Y} f\,(x,y) + \frac{L}{2}\ \| x \|^2 = arg \max_{y \in Y} f(x,y).$$

Let $y^*(x) = arg \max_{y \in Y} f(x,y)$, then

$$g(x') = f\big(x',y^*(x)\big) \geq^{(a)} f\big(x,y^*(x)\big) + \langle u_{x,y^*(x)}, x' - x\ \rangle - \frac{L}{2}\ \| x' - x \|^2,$$

$$\Rightarrow^{(b)}\qquad g(x') \geq g(x) + \langle v_x, x' - x\ \rangle - \frac{L}{2}\ \| x' - x \|^2,$$

where (a) uses L−weak convexity of $f(.,y)$, and (b) uses (9) and $v_x \in\ \partial g(x)$. So g is L−weakly convex in x and $\partial g(x)$ is defined in (9) $\forall\ x\ \in\ X$.

6 Conclusion:

In this research, we study more general case of weak−convexity in optimization problem than in the previous works s.t. the function f depends on its ν−Hölder−continuous gradient to get weak−convex function g. Therefore, we have different level of smoothness of f related to $\nu \in (0,1]$. So maybe we can design new faster converging schemes for solving minimax problems or any other problems.

7 References:

[1] Berger, O. J. 2013, *Statistical decision theory and Bayesian analysis*, Springer Science & Business Media.

[2] Bertsekas, D. P. 2014, *Constrained optimization and Lagrange multiplier methods*, Academic press, The United States of America.

[3] Bertsekas, D. P. 2009, *Convex optimization Theory*, Belmont, Massachusetts.

[4] Boyd, S.; Vandenberghe, L. 2004, *Convex Optimization*, ISBN 978-0-521-83378-3, Retrieved 12 Apr 2021, Cambridge University Press, United Kingdom.

[5] Devolder, O.; Glineur, F.; Nesterov, Y. 2013, *First-Order Methods of Smooth Convex Optimization with Inexact Oracle*, Mathematical Programming. DOI: 10.1007/s10107-013-0677-5, Springer-Verlag Berlin Heidelberg and Mathematical Optimization Society.

[6] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A. and Bengio, Y. 2014, *Generative adversarial nets*, Advances in neural information processing systems, Montreal, Canada.

[7] Jin, C.; Netrapalli, P. and Jordan, M. 2019, "*Minmax Optimization: Stable Limit Points of Gradient Descent Ascent are Locally Optimal*", University of California, Berkeley.

[8] Kinderlehrer, D. and Stampacchia, G. 1980, *An introduction to variational inequalities and their applications*, ISBN: 9780080874043, Siam, New York, Vol. 312.

[9] Kruger, A. Ya. 2003, *On Fréchet Subdifferentials*, Springer, Journal of Mathematical Sciences, Vol.116.

[10] Luo, Z. Q. and Yu, W. 2006, *An Introduction to Convex Optimization for Communications and Signal Processing*, IEEE, Vol.24.

[11] Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D. and Vladu, A. 2017, *Towards deep learning models resistant to adversarial attacks*, Cornell university, arXiv:1706.06083

[12] Myerson, B. R. 1997, *Game theory*, ISBN: 9780674341166, Harvard university press, Cambridge, Massachusetts, London, England.

[13] Rafique, H.; Liu, M.; Lin, Q. and Yang, T. 2018, *Non-convex min-max optimization: Provable algorithms and applications in machine learning*, Cornell university, arXiv:1810.02060.

[14] Rockafellar, T. 1970, *Convex Analysis*, New Jersey: Princeton University Press, USA.

[15] Schmit, L.A.; Fleury, C. 1980: *Structural synthesis by combining approximation concepts and dual methods*, American Institute of Aeronautics and Astronautics, Vol.18.

[16] Thekumparampil, K. K.; Jain, P.; Netrapalli, P.; Oh, S. 2019, *Efficient Algorithms for Smooth Minimax Optimization*, arXiv:1907.01543, 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.