Linear convergence of mirror descent via Bregman distance under more general conditions

Dr. Yamar Hamwi*

(Received 26 / 11 / 2025. Accepted 29 / 3 / 2026)

□ ABSTRACT □

The mirror gradient algorithm is a first-order method based on the Bregman distance for convex optimization. Traditional analyses typically assume Lipschitz continuous gradients and strongly convex Bregman kernels. Recent works introduced relative smoothness to relax the Lipschitz condition. However, convergence guarantees under the absence of both assumptions remain limited. This paper establishes global convergence to a minimizer and linear convergence to the optimal solution set using only relative smoothness, without requiring strong convexity of the kernel or Lipschitz continuity of the objective gradient. These results broaden the applicability of mirror descent to a wider class of problems.

Keywords: Convex optimization, Bregman distance, Distance kernel, Bregman function Relative smoothness, Mirror descent, Linear rate of convergence, Global convergence.

Copyright **:**Latakia University Journal (Formerly Tishreen) -Syria, The authors retain the copyright under a CC BY-NC-SA 04

\

* **Assistant Professor, Department of Informatics Engineering, Faculty of Engineering , Manara University Lattakia, Syria.** Yamar.Hamwi@manara.edu.sy

التقارب الخطي لخوارزمية انحدار المرآة وفق مفهوم مسافة بريغمان ضمن شروط أكثر عمومية

د. يمار حموي *

(تاريخ الإيداع26 / 11 / 2025. قبل للنشر في 29 / 3 / 2026)

□ ملخّص □

خوارزمية انحدار المرآة هي طريقة من الدرجة الأولى تعتمد على مسافة بريغمان لحل مسائل الأمثلة الأصغرية المحدبة. تفترض التحليلات التقليدية عادةً تدرج دالة الهدف ليبشز مستمر ونواة مسافة بريغمان محدبة بقوة. وقد أدخلت الدراسات الحديثة مفهوم السلاسة النسبية لتخفيف شرط ليبشز. ومع ذلك لا تزال هناك فجوة تتمثل في إمكانية تحقق تقارب الخوارزمية لمسائل الأمثلة المحدبة في غياب هذين الشرطين التقليديين الصارمين. تهدف هذه المقالة لإثبات التقارب العالمي الى القيمة الصغرى ونُبيّن أن الخوارزمية متقاربة خطيًا إلى مجموعة الحلول المثلى بالاعتماد على مفهوم السلاسة النسبية ضمن شروط أكثر عمومية ودون اشتراط التحدب القوي لنواة بريغمان أو تدرج تابع الهدف ليبشز مستمر مما يعزز من جدواها في التطبيقات العملية ويفتح آفاقًا جديدة في حل مسائل الأمثليات بكفاءة أعلى.

الكلمات المفتاحية: الأمثليات المحدبة، مسافة بريغمان، نواة المسافة، تابع بريغمان، السلاسة النسبية، انحدار المرآة، معدل التقارب الخطي، التقارب العالمي.

حقوق النشر : مجلة جامعة اللاذقية (تشرين سابقاً)– سوريا، يحتفظ المؤلفون بحقوق النشر بموجب الترخيص CC BY-NC-SA 04

* **مدرسة– قسم هندسة المعلوماتية –كلية الهندسة– جامعة المنارة – اللاذقية – سوريا** Yamar.Hamwi@manara.edu.sy

مقدمة:

تُعد مسائل الأمثليات المحدبة ركيزة أساسية في العديد من مجالات العلوم والهندسة، بدءاً من التعلم الآلي وتحليل البيانات الضخمة، مرورًا بمعالجة الإشارات والرؤية الحاسوبية، وصولًا إلى النمذجة العددية في الفيزياء والهندسة. نهتم في هذه المقالة بدراسة مسألة القيمة الدنيا المقيدة المعرّفة بالشكل:

$$(P): \quad inf\{f(x) \ : x \in C\}$$

حيث f تابع محدّب والمجموعة C مجموعة جزئية من $\mathbb{R}^n$ محدبة مغلقة غير خالية.

تحظى الطرق من الدرجة الأولى اهتمام كبير بسبب بساطتها وتكلفتها المنخفضة لكل تكرار وتعد طريقة مسقط الانحدار التدرجي (projected gradient descent) طريقة كلاسيكية لحل(P) ، ويتألف من تصغير التقريبات التربيعية لـ f على التوالي، وتأخذ الشكل:

$$x_{k+1} = \underset{u\in C}{arg\,min}\left\{f(x_k) + \langle \nabla f(x_k), u - x_k\rangle + \frac{1}{2\lambda}\|u, x_k\|^2\right\} (2)$$

عمّمت الطريقة المعرّفة في(2) بطريقة انحدار المرآة (Mirror descent) وذلك باستبدال المسافة الإقليدية $\frac{1}{2}\|.\|^2$ في (2) بمسافة بريغمان $D_h(.,.)$(المحددة في العلاقة (4) أدناه)، وتأخذ العلاقة التكرارية الشكل الآتي:

$$x_{k+1} = \underset{u\in C}{arg\,min}\left\{f(x_k) + \langle \nabla f(x_k), u - x_k\rangle + \frac{1}{\lambda}D_h\big(u, x_{k,}\big)\right\} (3)$$

حيث $\lambda > 0$ و h تابع نواة المسافة ممّا يسمح اختيارh بتصميم خوارزميات من الدرجة الأولى تتكيف مع هندسة مجموعة القيود أو دالة الهدف.

اقتُرحت طريقة انحدار المرآة في [20، 21] وقدّم تحليلًا مبسطًا لها في [6]. تُعتبر هذه الطريقة امتدادًا غير إقليدي لطريقة مسقط الانحدار التدرج الكلاسيكية، كما تُطبق على مسائل الأمثلة في فضاءات باناخ [15]. واقتُرح امتداد الطريقة للمسائل المقيدة في [20، 5]. تُستخدم هذه الطريقة في العديد من التطبيقات [7,8,19,27]، تم تحسين الطريقة وتوسيعها من وجهات نظر مختلفة. تتطلب الافتراضات المعتادة والمعيارية في العديد من الأعمال السابقة لتحديد معدل التقارب وتعقيد التكرار لطريقة انحدار المرآة باستخدام مسافة بريغمان كلاً من تدرج تابع الهدف ليبشز مستمر والتحدب القوي لنواة المسافة كما في [1,16]. فقدّم في [3] مفهوم السلاسة النسبية (LC) الذي يحل محل خاصية استمرارية ليبشز المعتادة حيث يُستبدل التقريب التربيعي العلوي المعتاد للتابع بمقياس تقارب أعم، كما قدم في [18] شرط التحدب القوي النسبي (RSC) كركيزة أساسية لضمان معدل التقارب الخطي، بينما قدمت أعمال أخرى مثل [28, 26] أطراً موحدة تربط بين خوارزمية انحدارا المرآة وطرق النقطة المقربة في فضاءات باناخ. على الرغم من المتانة النظرية لهذه الأطر، إلا أن خوارزميتهم وتعقيد التقارب يعتمدان على مقياس تناظر لمسافة بريغمان وضمان التقارب الخطي ظل رهيناً بشرط (RSC) الذي يتطلب وجود انحناء بين أي نقطتين في نطاق الدالة. فهي افتراضات تبقى صارمة ولا تتحقق في طيف واسع من مسائل التعلم الآلة لاسيما في الأبعاد العالية مثل تدريب الشبكات العصبية العميقة أو الانحدار اللوجستي على بيانات غير قابلة للفصل. في مثل هذه الإعدادات، يكون التحدب القوي حتى بصيغته النسبية غائباً أو صعب التحقق عملياً، مما يحدّ من قابلية تطبيق النتائج للتقارب الخطي لانحدار المرآة.

نهتم في هذه المقالة بتخفيف هذه المتطلبات استناداً الى شرط (LC) من خلال تقديم شرط كافٍ يُعرف بـ شرط نمو بريغمان (Bregman Growth Condition). وتتمثل المساهمة الجوهرية في إثبات التقارب الخطي تحت افتراض أضعف لا يستلزم انحناءً في كافة الاتجاهات كما يفترض(RSC) ، بل يكفي وجود نمو أدنى للدالة

بالنسبة لمجموعة الحلول المثلى وهو يختلف قليلاً عن الدراسات السابقة حيث لا يتضمن معامل التناظر (Symmetry coefficient) الذي قدّم في [3]. مما يزيد من فائدة طرق التدرج في التطبيقات العملية ويفتح آفاقاً لنمذجة مسائل تعلم الآلة بمرونة رياضية أكبر

أهمية البحث وأهدافه:

يهدف البحث إلى دراسة تحليل التقارب لخوارزمية انحدار المرآة باعتباره امتدادًا غير إقليدي لانحدار التدرج وفق مفهوم مسافة بريغمان باستخدام السلاسة النسبية. وتقديم شرط نمو بريغمان الذي يؤدي إلى ضمانات تقارب تحت افتراضات أكثر عمومية لحل مسألة أصغريات. مما يسمح بإثبات التقارب العالمي لنقطة حرجة والتقارب خطي والتي لا يتطلب تحدب قوي نسبي أو تدرج تابع ليبشز مستمر. وتكمن أهمية البحث بتطبيق أوسع لمسائل الأمثليات غير الملساء حيث تفشل الافتراضات التقليدية. مما يقدم أساساً نظرياً أكثر ملاءمة وكفاءة هندسية لتدريب نماذج الذكاء الاصطناعي واسعة النطاق.

طرائق البحث ومواده:

نقدّم بعض التعاريف والمفاهيم الأساسية المتعلقة بالتحليل المحدّب ومفهوم السلاسة النسبية الذي يعتبر أداة أساسية لدراسة خوارزمية مرآة وكما نقدّم تعريف مسافة بريغمان وأهم خواصها التي تساعدنا في معالجة مسألتنا.

تعاريف ومفاهيم أساسية:

نبدأ ببعض التعاريف والمفاهيم الأساسية التي تستخدم في التحليل المحدب الذي يلعب دوراً مهم في نظرية الأمثليات ومن أجل تفاصيل أكثر يمكن العودة إلى[1,22,23,24] .

– نرمز للصاقه مجموعة $C \subseteq \mathbb{R}^n$ بـ $\overline{C}$ وداخليتها بـ لتكن $int\, C$

– لتكن $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$، يقال عن f إنها دالة خاصة (proper) إذا كان المجال الفعلي $domf \neq \emptyset$ **حيث** $domf = \{x \in \mathbb{R}^n : f(x) < +\infty\}$.

مسافة بريغمان:

قدّمها بريغمان في [11] ودرست من قبل العديد من الباحثين وخلال السنوات الماضية تمّ استخدامها في مجالات مختلفة مثل تعلّم الآلة (machine learning) ومعالجة الصورة (image processing) وتجميع البيانات (data clustering) والمعادلات التفاضلية الجزئية غير الخطية والعديد من المجالات الأخرى [2,5,12,17]. الخطوة الأولى في تعريف مسافة بريغمان هي اختيار نواة المسافة h (distance kernel).

لتكن C مجموعة جزئية محدبة مغلقة غير خالية من $\mathbb{R}^n$. يقال عن $h: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ إنه تابع **نواة المسافة** على C إذا تحقق:

i. $C = \overline{domh}$
ii. h محدبة ونصف مستمرة من الأدنى وخاصة.
iii. h قابل للاشتقاق باستمرار ومحدبة تماماً على $int\, domh \neq \emptyset$.

تعرّف **مسافة بريغمان (Bregman distance)** $D_h: domh \times int\ domh \to [0, +\infty[$ بالعلاقة الآتية:

$$D_h(x,y) = h(x) - h(y) - \langle \nabla h(y), x - y \rangle. \quad (4)$$

لاحظ أن D_h تابع حقيقي غير سالب لا يحقق متراجحة المثلث فهو ليس مسافة مترية. وإذا كان التابع $h(x) = \|x\|^2$ فإنّ $D_h(x,y) = \|x-y\|^2$، أي أنّ النظيم التربيعي حالة خاصة من مسافة بريغمان.

نعرض بعض خواص مسافة بريغمان في $\mathbb{R}^n$ من خلال التمهيديتين الآتيتين.

تمهيدية 1: [4]

ليكن $h, h_1, h_2: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ تابع محدب تماماً، عندئذٍ:

i. $D_h(x,y) \geq 0$ لكل $x \in domh$ و $y \in int\ domh$،

ii. $D_h(x,y) = 0$ إذا وفقط إذا كان $x = y$.

تمهيدية 2: [27]

إذا كان $h, h_1, h_2: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$، عندئذٍ:

i. (خاصة النقاط الثلاث) لكل $x \in domh$ و $y \in int\ domh$ ، فإنّ:

$$D_h(x,z) = D_h(x,y) + D_h(y,z) + \langle \nabla h(y) - \nabla h(z), x - y \rangle.$$

ii. (خاصة الخطية) لكل $\alpha, \beta \in \mathbb{R}$ و $x, y \in domh_1 \cap domh_2$ فإنّ:

$$D_{\alpha h_1 + \beta h_2}(x,y) = \alpha D_{h_1}(x,y) + \beta D_{h_2}(x,y).$$

مفهوم السلاسة النسبية (Relative smoothness)

نقدَم في هذا القسم مفهوم السلاسة النسبية والمعروف أيضًا باسم خاصية التكيف السلس (smooth adaptable property) ، أو حالة تشبه حالة ليبشز (Lipschitz−like condition) والذي يحقق مواءمة بين نواة بريغمان وهندسة تابع الهدف بما يحقق توافقاً بين مسافة بريغمان وبنية الفضاء الذي يُعرّف عليه التابع. تمّ تقديمه أول مرة في [3]، ويرمز له ب (LC).

لنفترض أن h تابع نواة المسافة على C ، و f تابع بحيث $domh \subset domf$. يقال إنَ f ملساء بالنسبة إلى h إذا وجد ثابت $L > 0$، **بحيث**

$(LC) \quad int\ domh$ محدبة على $Lh - f$

يكافئ شرط LC مبرهنة الانحدار (descent lemma) في [3]، ولكن هناك افترض أن f محدبة. ولكن بيّن في [10،25] يمكن أن تكون f غير محدبة،` أي تحدب f لا يلعب أي دور في شرط LC. لندرس بعض خواصه من خلال التمهيديات التالية:

تمهيدية 3: [18]

بفرض زوج من التوابع (f,h) يحقق شرط (LC) و $L > 0$ ، العبارات التالية متكافئة:

i. $f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + LD_h(x,y)$ لكل $x, y \in int\ domh$

ii. $D_f(x,y) \leq LD_h(x,y)$ لكل $x, y \in int\ domh$

iii. $D_{Lh-f}(x,y) \geq 0$ لكل $x, y \in int\ domh$

iv. $\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L\big(D_h(x,y) + D_h(y,x)\big)$ لكل $x \in int\ domh$

v. إذا كان كل من f, h قابل للاشتقاق مرتين عندئذٍ لكل $x \in int\ domh$ فإنّ

$$\nabla^2 h(x) - \nabla^2 f(x) \geq 0$$

من الواضح أنه عندما يكون f محدبًا، فإن $Lh+f$ تابع محدب وضوحا. تسمح خواص D_h باستنتاج نتيجة أخرى مفيدة وأكثر عمومية تسمى تمهيدية الانحدار بثلاث نقط الآتية.

تمهيدية 4: [10]

بفرض h تابع نواة المسافة على C ، و f تابع بحيث $domh \subset domf$ و $L>0$، عندئذٍ الزوج (f,h) يحقق شرط (LC) إذا وفقط إذا كان لكل $x,y,z \in int\ domh$

$$f(x) \le f(y) + \langle \nabla f(z), x-y\rangle + LD_h(x,z) - D_f(y,z) \quad (5)$$

وإذا كان f تابع محدب يمكننا إسقاط الحد الأخير في المتراجحة (5) ، ويكون:

$$f(x) \le f(y) + \langle \nabla f(z), x-y\rangle + LD_h(x,y) \quad (6)$$

عادةً ما تتطلب الطرق من الدرجة الأولى المبنية على مسافات بريغمان بأن يكون تدرج f ليبشز مستمر والنواة h محدبة بقوة. لكن من خلال النتيجة البسيطة التالية تُظهر أن هذين الافتراضين حالة خاصة من شرط (LC).

تمهيدية 5: [25]

بفرض h نواة المسافة محدبة بقوة، و f تدرجه ليبشز مستمر على $L>0$، عندئذٍ الزوج (f,h) يحقق شرط (LC) أي أن التابع $Lh-f$ محدب على $x \in int\ domh$.

والعكس غير صحيح أي إذا كان الزوج (f,h) يحقق شرط (LC) فليس بالضرورة أن يكون التابع f تدرجه ليبشز مستمر على $int\ domh$ كما هو مبين بالمثال التالي.

مثال: ليكن لدينا التابعان h,f معرفان بالعلاقة التالية:

$$h(x) = -5lnx, domh =]0,+\infty[$$
$$f(x) = -3x^{-2}, \qquad domf = \mathbb{R}\backslash\{0\}$$

نلاحظ أنّ $domh \subset domf$ وتعطى قاعدة ربط التابع $Lh-f$ بالعلاقة:

$$(Lh-f)(x) = -5L.lnx + 3x^{-2}.$$

والتابع Lh - f محدب على $int\ domh$ لأن لكل $x \in int\ domh$ و $L>0$ يكون

$$(Lh-f)''(x) = 5Lx^{-2} + 18x^{-4} > 0.$$

أي الزوج(f,h) يحقق شرط (LC)، ولكن تدرج التابع f ليس ليبشز مستمر على $int\ domh$ لأن $f'(x) = 6x^{-3}$ لكل $x \in int\ domh$ و $f''(x) = -18x^{-4}$ ليس محدود على $int\ domh$.

المسألة وخوارزمية انحدار المرآة

هدفنا هو حل مسألة الأصغرية الدنيا(P): $inf\{f(x):\ x \in C\}$ وفقًا للافتراض القائم الآتي:

الفرض A:

i. C هي مجموعة محدبة مغلقة غير خالية من $\mathbb{R}^n$.

ii. الدالة $h: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ تابع نواة المسافة على C

iii. التابع $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ خاص ونصف مستمر من الأدنى حيث $domh \subset domf$.

iv. $-\infty < inf\{f(x):\ x \in C = \overline{domh}\} = inf\{f(x):\ x \in domh\}$.

v. لكل $\lambda > 0$ و $x \in int\ domh$ و $p \in \mathbb{R}^n$ للمسألة:

$$\min_{u \in C}\left\{\langle p, u-x\rangle + \frac{1}{\lambda}D_h(u,x)\right\}$$

حل أصغري وحيد في$domh$.

الشرط (v) هو شرط يضمن أنّ طريقة انحدار المرآة معرفة جيدا بالإضافة إلى هذه الافتراضات، فإن الخاصية الأساسية التي نحتاجها لتطبيق طريقة انحدار المرآة هو شرط (LC).

- نرمز إلى الزوج (f,h) ينتمي الى $R_L(C)$ إذا كان:
  i. التابعان f و h يحققان الفرض A
  ii. التابع $Lh-f$ محدب على C.

خوارزمية انحدار المرآة

تسمح الافتراضات السابقة بتحديد خوارزمية انحدار المرآة لحل المسألة (P).

الخوارزمية:

المدخلات: الزوج (f,h) ينتمي الى$R_L(C)$ و $x_0 \in int\ domh$ قيمة ابتدائية وحجم خطوة مختار $0<\lambda<\frac{1}{L}$.

الخطوة العامة: من أجل $k=1,2,\ldots$،

$$x_{k+1}=\underset{u\in C}{arg\,min}\left\{\langle\nabla f(x_k),u-x_k\rangle+\frac{1}{\lambda}D_h(u,x_k)\right\}(7)$$

يتطلب عادة تحليل التقارب لطرق الدرجة الأولى المستندة إلى مسافة بريغمان بأن يكون تدرج تابع الهدف ليبشز مستمر وتابع نواة المسافة محدب بقوة، وهذه الافتراضات قوية في التطبيقات العملية. نهتم في هذه المقالة بتخفيف هذه المتطلبات لتحديد معدل التقارب خوارزمية انحدار المرآة بالإضافة لذلك حيث لا يتضمن معامل التناظر ونواة بريغمان دالة ليجندر.

النتائج والمناقشة:

ندرس في هذا القسم تحليل التقارب ونحلل خواص الخوارزمية المعطاة بالعلاقة التكرارية (7) تحت فرضيات معرفة جيدا دون أن تضمن معامل التناظر. ونعرضه في تسلسل منهجي يبدأ بإثبات خاصية الانحدار الأساسية ثمّ نبين الاطراد وصولاً الى إثبات التقارب العالمي وأخيراً التقارب الخطي تحت شرط نمو بريغمان. والمبرهنة التالية هي المفتاح الرئيسي لدراسة نتائج التقارب.

مبرهنة 6:

بفرض الزوج (f,h) ينتمي الى$R_L(C)$ فلكل $x\in int\ domh$ ولكل $\tilde{x}\in int\ domh$ و

$$\tilde{x}\in T_\lambda(x)=\underset{u\in C}{arg\,min}\left\{\langle\nabla f(u-x\rangle+\frac{1}{\lambda}D_h(u,x)\right\}(\lambda>0)$$

عندئذٍ:

i. $\lambda f(\tilde{x})\le\lambda f(x)-(1-\lambda L)D_h(\tilde{x},x).(8)$
ii. $\lambda f(\tilde{x})\le\lambda f(x)+\lambda\langle\nabla f(x),_{,}u-x_{,}\rangle-(1-\lambda L)D_h(\tilde{x},x)+D_h(u,x_{,})-D_h(u,\tilde{x}_{,})(9)$
iii. $\lambda f(\tilde{x})\le\lambda f(u)+D_h(u,x)-D_h(u,\tilde{x})-(1-\lambda L)D_h(\tilde{x},x)-\lambda D_f(u,x)\ (10)$

وكنتيجة لذلك إذا كانت f محدبة تماما فإن

$$\lambda f(\tilde{x})\le\lambda f(u)+D_h(u,x)-D_h(u,\tilde{x})-(1-\lambda L)D_h(\tilde{x},x)\ (11)$$

البرهان:

i. حسب الفرض لكل $u \in C$ لدينا:

$$\langle \nabla f(x), \tilde{x} \rangle + \frac{1}{\lambda} D_h(\tilde{x}, x) \leq \langle \nabla f(x), u \rangle + \frac{1}{\lambda} D_h(u, x).$$

بوضع $u = \tilde{x}$، نحصل على

$$\langle \nabla f(x), \tilde{x} \rangle + \frac{1}{\lambda} D_h(\tilde{x}, x) \leq \langle \nabla f(x), x \rangle.$$

وحسب الجزئية (i) من التمهيدية 3، يكون:

$$f(\tilde{x}) \leq f(x) + \langle \nabla f(x), \tilde{x} - x \rangle + L D_h(\tilde{x}, x)$$

وبجمع المتراجحتين السابقتين، يكون $\lambda f(\tilde{x}) \leq \lambda f(x) - (1 - L\lambda) D_h(\tilde{x}, x)$.

ii. لكل $\lambda > 0$ لدينا فرضاً $T_\lambda(x) =$ $arg\ min_{\ u \in C} \left\{ f(x) + \langle \nabla f(x), u - x \rangle + \frac{1}{\lambda} D_h(u, x) \right\}$ بحذف الحدود الثابتة، يكون

$$T_\lambda(x) = \underset{u \in C}{arg\ min} \left\{ \langle \nabla f(x), u \rangle + \frac{1}{\lambda} D_h(u, x) \right\} (12)$$

أي $0 \in \partial \left\{ \langle \nabla f(x), . \rangle + \frac{1}{\lambda} D_h(., x) \right\} (\tilde{x})$ ومنه

$$0 \in \partial \{ \langle \nabla f(x), . \rangle \} (\tilde{x}) + \frac{1}{\lambda} [\nabla h(\tilde{x}) - \nabla h(x)].$$

أي $z = \frac{1}{\lambda} \left(\nabla h(x) - \nabla h(\tilde{x}) \right) \in \partial \{ \langle \nabla f(x), . \rangle \} (\tilde{x})$ وبالاعتماد على مفهوم تحت التفاضل، نجد أنّ:

$$\langle \nabla f(x), u \rangle - \langle \nabla f(x), \tilde{x} \rangle \geq \langle u - \tilde{x}, z \rangle.$$

أي $$\lambda \langle \nabla f(x), \tilde{x} - u \rangle \leq \langle \tilde{x} - u, \nabla h(\tilde{x}) - \nabla h(x) \rangle.$$

وحسب الجزئية (i) من التمهيدية 2، يكون:

$$\lambda \langle \nabla f(x), \tilde{x} - u \rangle \leq D_h(u, x) - D_h(u, \tilde{x}) - D_h(\tilde{x}, x). (13)$$

من الجزئية (i) من التمهيدية 3، يكون:

$$f(\tilde{x}) \leq f(x) + \langle \nabla f(x), \tilde{x} - x \rangle + L D_h(\tilde{x}, x)$$

وبجمع المتراجحتين السابقتين، نحصل على

$$\lambda f(\tilde{x}) \leq \lambda f(x) + \lambda \langle \nabla f(x), u - x \rangle - (1 - L\lambda) D_h(\tilde{x}, x) + D_h(u, x) - D_h(u, \tilde{x})$$

iii. من التمهيدية 4، يكون:

$$f(\tilde{x}) \leq f(u) + \langle \nabla f(x), \tilde{x} - u \rangle + L D_h(\tilde{x}, x) - D_f(u, x) (14)$$

بضرب طرفي العلاقة (14) بـ λ وجمعها الى المتراجحة (13) ، يكون

$$\lambda f(\tilde{x}) \leq \lambda f(u) + D_h(u, x) - D_h(u, \tilde{x}) + (1 - L\lambda) D_h(\tilde{x}, x) - \lambda D_f(u, x).$$

وكنتيجة لذلك إذا كانت f محدبة تماما فحسب الجزئية (i) من التمهيدية 1 يكون $D_f(u, x) \geq 0$ ومنه

$$\lambda f(\tilde{x}) \leq \lambda f(u) + D_h(u, x) - D_h(u, \tilde{x}) + (1 - L\lambda) D_h(\tilde{x}, x).$$

وبذلك يتم المطلوب. ■

تبين المبرهنة 6 أن الشرط $\lambda \in \left]0, \frac{1}{L}\right[$ يضمن انخفاضًا كافيًا في قيمة دالة الهدف. هذه الحقيقة تعطي النتيجة التالية.

تمهيدية 7:

ليكن الزوج (f,h) ينتمي الى $R_L(C)$ و $\{x_n\}_{n\in\mathbb{N}}$متتالية معرّفة بالعلاقة(7)، حيث $\lambda \in \left]0,\frac{1}{L}\right[$ عندئذٍ:

i. المتتالية $\{f(x_k)\}_{k\in\mathbb{N}}$متناقصة.

ii. $\sum_{k=1}^{\infty} D_h(x_k,x_{k-1}) < \infty$ وكنتيجة لذلك المتتالية $\{D_h(x_k,x_{k-1})\}_{k\in\mathbb{N}}$ تقترب من الصفر.

iii. $\min_{1\le k\le n} D_h(x_k,x_{k-1}) \le \frac{\lambda[f(x_0)-f^*]}{(1-\lambda L)}$حيث $f^* = inf\ \{f(x):\ \ x\in C\}$

برهان.

i. حسب تعريف المتتالية x_k، وباستبدال x بـx_{k-1}، و $\tilde{x}$ بـ x_k، في العلاقة (8)، نحصل على:

$$\lambda f(x_k) \le \lambda f(x_{k-1}) - (1-\lambda L)D_h(x_k,x_{k-1}). (15)$$

بما أنّ $\lambda \in \left]0,\frac{1}{L}\right[$ وحسب خواص مسافة بريغمان$D_h(x_k,x_{k-1}) \ge 0$ يكون $f(x_k) \le f(x_{k-1})$، والمتتالية $\{f(x_k)\}_{k\in\mathbb{N}}$ متناقصة.

ii. بأخذ المجموع لـ k من 1 إلى n لطرفي العلاقة (15)، نحصل على:

$$\sum_{k=1}^{n} D_h(x_k,x_{k-1}) \le \frac{\lambda[f(x_0)-f(x_n)]}{(1-\lambda L)} \le \frac{\lambda[f(x_0)-f^*]}{(1-\lambda L)} (16)$$

حيث $f^* = inf\ \{f(x):\ \ x\in C\} \succ \infty$ وبجعل $n\to\infty$ نحصل على $\sum_{k=1}^{\infty} D_h(x_k,x_{k-1}) < \infty$ ومنه المتتالية $\{D_h(x_k,x_{k-1})\}_{k\in\mathbb{N}}$ تقترب إلى الصفر.

iii. من العلاقة (16)، نجد أنّ:

$$n \min_{1\le k\le n} D_h(x_k,x_{k-1}) \le \sum_{k=1}^{n} D_h(x_k,x_{k-1}) \le \frac{\lambda[f(x_0)-f^*]}{(1-\lambda L)}$$

وبقسمة طرفي العلاقة السابقة علىn، يكون:

$$\min_{1\le k\le n} D_h(x_k,x_{k-1}) \le \frac{\lambda[f(x_0)-f^*]}{n(1-\lambda L)}$$

وبذلك يتم المطلوب. ■

تعد هذه المبرهنة نتيجة تمهيدية لتقارب قيم الدالة إلى الحد الأدنى، مما يمهد الطريق لإثبات التقارب العالمي في النتائج اللاحقة. وقد أتاح ذلك في المبرهنة الآتية إثبات معدل تقارب خوارزمية انحدار المرآة $O(1\backslash n)$ الذي يمثل تمديداً لمعدل تقارب انحدار التدرج الإقليدي دون الاعتماد على معامل التناظر. وكما تثبت المبرهنة أنّ المتتالية $f(x_k)$ تتقارب إلى القيمة المثلى لمسألة الأمثلة الأصغرية P ، حتى وإن لم تكن المتتالية x_k ليست متقاربة بالضرورة.

مبرهنة 8:

ليكن الزوج (f,h) ينتمي الى $R_L(C)$ و f محدبة تماماً و $\{x_n\}_{n\in ¥}$ متتالية معرّفة بالعلاقة (7)، حيث$L>0$ عندئذٍ:

i. من أجل $\lambda \in \left]0,\frac{1}{L}\right[$ ولكل $n\ge 1$ يكون $f(x_n)-f(u) \le \frac{1}{n\lambda} D_h(u,x_0)$.

ii. $\lim_{n\to\infty} f(x_n) = \inf_{x\in C} f(x)$.

البرهان.

i. بما أن $x_k \in T_\lambda(x_{k-1})$، وحسب الجزئية (iii) من مبرهنة 6 يكون:

$$\lambda f(x_k) \le \lambda f(u) + D_h(u, x_{k-1}) - D_h(u, x_k) - (1 - \lambda L) D_h(x_k, x_{k-1})$$

وبإعادة الترتيب نجد أنّ

$$f(x_k) - f(u) \le +\frac{1}{\lambda}[D_h(u, x_{k-1}) - D_h(u, x_k)] - \frac{(1 - \lambda L)}{\lambda} D_h(x_k, x_{k-1})$$

وحسب خواص مسافة بريغمان $D_h(.,.) > 0$ و $0 < \lambda L < 1$ يكون $\frac{(1-L\lambda)}{\lambda} D_h(x_k, x_{k-1}) \ge 0$، ومنه

$$f(x_k) - f(u) \le \frac{1}{\lambda}[D_h(u, x_{k-1}) - D_h(u, x_k)] \forall k \ge 1$$

بأخذ المجموع لـ k من 1 إلى n لطرفي العلاقة السابقة، نحصل على:

$$\sum_{k=1}^{n}[f(x_k) - f(u)] \le \frac{1}{\lambda}\sum_{k=1}^{n}[D_h(u, x_{k-1}) - D_h(u, x_k)]$$

نلاحظ الطرف الأيمن من المتراجحة هو مجموع تلسكوبي حيث تختزل الحدود المتتالية ويكون:

$$\sum_{k=1}^{n}[D_h(u, x_{k-1}) - D_h(u, x_k)] = D_h(u, x_0) - D_h(u, x_n)$$

ومنه المتراجحة تؤول إلى:

$$\sum_{k=1}^{n}[f(x_k) - f(u)] \le \frac{1}{\lambda}[D_h(u, x_0) - D_h(u, x_n)]$$

وحسب خواص مسافة بريغمان يكون $D_h(u, x_n) \ge 0$ نستنتج أنّ:

$$\sum_{k=1}^{n}[f(x_k) - f(u)] \le \frac{1}{\lambda}[D_h(u, x_0) - D_h(u, x_n)] \le \frac{1}{\lambda} D_h(u, x_0). \quad (17)$$

من جهة ثانية بما أنّ المتتالية $\{f(x_k)\}_{k\in\mathbb{N}}$ متناقصة حسب الجزئية (i) من تمهيدية 7 فالمتتالية $v_k = f(x_k) - f(u)$ متناقصة، ومنه $v_n \le \frac{1}{n}\sum_{k=1}^{n} v_k$ أي

$$f(x_n) - f(u) \le \frac{1}{n}\sum_{k=1}^{n}[f(x_k) - f(u)]$$

وحسب العلاقة (17) ، يكون

$$f(x_n) - f(u) \le \frac{1}{n\lambda} D_h(u, x_0), \quad \forall u \in domh \quad (18)$$

ii. ليكن x^* حل أصغري لـ f أي $f\left(x^*\right) = \inf_{x\in C} f(x)$، من جهة أولى حسب الجزئية (iv) من الفرض A يكون $f\left(x^*\right) \succ \infty$، وبما أنّ المتتالية $\{f(x_k)\}_{k\in\mathbb{N}}$ متناقصة فحسب الجزئية (i) من تمهيدية 7 يكون:

$$\lim_{n\to\infty} f(x_n) \ge f(x^*) = \inf_{x\in C} f(x) \succ \infty.$$

ومن جهة ثانية بجعل n تسعى إلى $+\infty$ لطرفي العلاقة (18)، يكون $\lim_{n\to\infty} f(x_n) \le f(u)$ لكل $u \in domh$، أي $\lim_{n\to\infty} f(x_n) \le f(x^*)$ ومنه $\lim_{n\to\infty} f(x_n) = f\left(x^*\right) = \inf_{x\in C} f(x)$.

وبذلك يتم المطلوب. ■

ملاحظة: ليكن $x^* \in domh$ حل أصغري لـ f، نضع $u = x^*$ في العلاقة (18) فنحصل على

$$f(x_n) - f(x^*) \le \frac{1}{n\lambda} D_h(x^*, x_0).$$

قدمت هذه المبرهنة نتيجة تمهيدية تثبت معدل تقارب $O(1\backslash n)$ ضمن إطار السلاسة النسبية وفي حين اعتمدت الدراسات السابقة (انظر [3، 18]) في اثباتها على وجود معامل تناظر، فإننا قدمنا إثبات في غياب هذا الشرط مما يمنح الخوارزمية مرونة أكبر. إلا أن الإسهام الجوهري لهذه المقالة يتجلى في المبرهنات اللاحقة التي تطرح تحليلاً متقدماً لمعدلات التقارب، بما في ذلك التقارب الخطي تحت شروط أكثر عمومية والتقارب العالمي في الحالات غير الملساء، وهو ما يشكل تطويراً مهماً للنتائج في هذا المجال.

التقارب العالمي

سنقدم تعريف تابع بريغمان من أجل اثبات التقارب العالمي. تمّ تعريف تابع بريغمان في [14] الذي يوفر بيئة مناسبة لدراسة مسافة بريغمان، وقد أثبت أن بعض الشروط تنتج من بعضها الآخر، لذلك قدّم الباحثون تعريفاً مكافئاً أكثر مرونة وفعالية.

تعريف 9: [12]

لتكن S مجموعة جزئية غير خالية من $\mathbb{R}^n$ مفتوحة ومحدبة. يقال إنّ $h: \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ تابع بريغمان على S إذا وفقط إذا تحققت الشروط الآتية:

a. h تابع مستمر ومحدب تماماً على $\overline{S}$.

b. h تابع قابل للاشتقاق على S.

c. لكل $\alpha \in \mathbb{R}^+, x \in \overline{S}$ فإن:

$R(x,\alpha) = y \in S; D_h(x,y) \leq \alpha$ محدودة.

d. إذا كانت $\{y_k\}_{k\in\mathbb{N}} \in S$ متقاربة حيث $\lim_{k\to\infty} y_k = y \in \overline{S}$ فإن

$$\lim_{k\to\infty} D_h(y, y_k) = 0$$

مثال: إذا كانت $S = \mathbb{R}^n_+$ و $h(x) = \sum_{i=1}^n \left(x_i - \sqrt{x_i}\right)^2$ فإن h تابع بريغمان، وبالتالي مسافة بريغمان تكون $D_h(x,y) = \sum_{i=1}^n \frac{1}{2\sqrt{y_i}} \left(\sqrt{x_i} - \sqrt{y_i}\right)^2$.

تمهيدية 10: [13]

لتكن S مجموعة جزئية غير خالية من $\mathbb{R}^n$ مفتوحة ومحدبة و $h: \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ دالة بريغمان على S عندئذ h تحقق الآتي:

إذا كانت $\{y_k\}_{k\in\mathbb{N}} \in S$ متقاربة حيث $\lim_{k\to\infty} y_k = y$ ، وإذا كانت المتتالية $\{x_k\}_{k\in\mathbb{N}} \in \overline{S}$ محدودة و $\lim_{k\to\infty} D_h(x_k, y_k) = 0$ فإنّ $\lim_{k\to\infty} x_k = y$.

تمهيدية 11 : [22]

إذا كانت x_n, y_n متتاليتين موجبتين من الأعداد الحقيقية وتحقق

i. $(\forall n \geq 0), x_{n+1} \leq x_n + y_n$.

ii. $\sum_{n=1}^{\infty} y_n < \infty$

عندئذٍ المتتالية x_n متقاربة.

تبين المبرهنة التالية التقارب العالمي للخوارزمية المعرفة بالعلاقة التكرارية (7) بفرض h تابع بريغمان دون فرض معامل التناظر.

مبرهنة 12:

ليكن الزوج (f,h) ينتمي الى $R_L(C)$ و h تابع بريغمان و fمحدبة تماماً ولتكن $\{x_k\}_{k\in\mathbb{N}}$ متتالية معرّفة بالعلاقة(7)، حيث $\lambda \in \left]0,\frac{1}{L}\right[$ عندئذٍ المتتالية x_kتتقارب من حلول المسألةP.

البرهان.

سيتم البرهان بثلاث خطوات

الخطوة الأولى: المتتالية x_k محدودة

الخطوة الثانية: سنبرهن أن كل نقاط التراكم للمتتالية تنتمي إلى مجموعة حلول المسألة.

الخطوة الثالثة: سنبرهن أن المتتالية متقاربة إلى حلول المسألة P.

الخطوة الأولى: من العلاقة (11)لكل $x \in domh = \overline{domh}$ يكون

$$D_h(x,x_k) \le D_h(x,x_{k-1}) - (1-\lambda L)D_h(x_k,x_{k-1}) - \lambda\big(f(x_k)-f(x)\big)$$

بوضع $x = \overline{x} \in \underset{u\in domh}{arg\,min} f(u)$ أي $f(\overline{x}) \le f(x_k)$ ومنه لكل kالمتراجحة السابقة تصبح:

$$D_h(\overline{x},x_k) \le D_h(\overline{x},x_{k-1}) - (1-\lambda L)D_h(x_k,x_{k-1}) \quad (19)$$

حسب الجزئية (ii) من تمهيدية7، يكون $\sum_{k=1}^{\infty} D_h(x_k,x_{k-1}) < \infty$ وكنتيجة لذلك المتتالية $\{D_h(x_k,x_{k-1})\}_{k\in\mathbb{N}}$ تقترب من الصفر. فحسب المتراجحة(19) وتمهيدية 11 فإنّ المتتالية $D_h(u,x_k)$ محدودة، أي يوجد$\sigma \ge 0$ حيث $D_h(u,x_k) \le \sigma$، بما أنّ h دالة بريغمان فإنّ المجموعة $E = \{y: D_h(u,y) \le \sigma\}$ محدودة، أي المتتالية$x_k \in E$ ، أي المتتالية x_k محدودة.

<u>الخطوة الثانية:</u>

بما أنّ (P) غير خالية ومتراصة وبرهنا في الخطوة الأولى أنّ المتتالية x_k محدودة فهي تملك نقطة تراكمx^* أي توجد متتالية جزئية x_{k_i} من x_k تتقارب من x^* وبالتالي يكون

$$inf\ f(x) = \lim_{k\to+\infty} f(x_{k_i}) = \liminf_{k\to+\infty} f(x_{k_i}). \quad (20)$$

وبما أنّ f نصف مستمرة من الأدنى، يكون:

$$f(x^*) \le \liminf_{k\to+\infty} f(x_{k_i}). \quad (21)$$

من العلاقتين (21) و (20) نجد أنّ $f(x^*) \le inf\ f(x)$ ومنه x^* حل أصغري لـ f.

<u>الخطوة الثالثة:</u>

لتكن x^* نقطة تراكم للمتتاليةx_k فحسب الخطوة الثانية x^* حل أصغري لـ f على$domh$ وبما أن h تابع بريغمان تكون المتتالية $D_h(x_k,x^*)$ متقاربه و $\lim_{k\to+\infty} D_h(x_k,x^*) = 0$، وحسب تمهيدية 10 يكون $\lim_{k\to+\infty} x_k = x^*$ ومنه $\overline{x}$ حل أصغري لـ f وبذلك يتم المطلوب. ■

تجدر الإشارة إلى أنَّ إثبات التقارب العالمي تحت شرط السلاسة النسبية قد طُرح في أعمال سابقة لمسألة أصغريات مركبة مثل [3، 18]، إلا أنَّ تلك النتائج اعتمدت بشكل جوهري على وجود معامل تناظر الذي يحكم العلاقة بين$D_h(x,y)$ و$D_h(y,x)$ بالإضافة إلى فرض قسرية دالة النواة h. في المقابل نُثبت في هذا العمل أن التقارب العالمي لخوارزمية انحدار المرآة تحت شرط السلاسة النسبية لا يتطلب هذه القيود الإضافية ويعتمد برهاننا على

الخصائص الجوهرية لدالة بريغمان المتمثلة في تعريف 9، مما يفتح الباب لتطبيقات أوسع في المسائل التي لا تحقق شروط التناظر التقليدية.

التقارب الخطي

تشير العديد من الأبحاث مؤخراً إلى أن افتراض التحدب القوي ليس شرطًا ضروريًا للحصول على تقارب خطي لطرق الدرجة الأولى فقدم في [17] شرط التحدب القوي النسبي (RSC) الذي يعمم مفهوم التحدب القوي التقليدي الآتي:

التابع f محدب بقوة نسبياً إذا وجد $\sigma > 0$ حيث

$$f(y) \geq f(x) + \langle \nabla f(y), u - x \rangle + \sigma D_h(y, x)). \quad \forall\, x, y \in int\ domh$$

أي $f - \sigma h$ محدبة حيث$\sigma > 0$. وقدم في [9] شرط النمو التربيعي الآتي:

التابع f يحقق شرط النمو التربيعي، إذا وجد $v > 0$ حيث

$$f(x) - f^* \geq \frac{v}{2} \inf_{y \in X} \|x - y\|^2, \forall x \in domf$$

حيث X هي مجموعة الحلول المثلى للمسألة (1) وf^* هي قيمة دالة الهدف المثلى.

نقدم شرط كافٍ وأكثر مرونة يُعرف بـ شرط نمو بريغمان (Bregman Growth Condition) لإثبات التقارب الخطي للخوارزمية ضمن شروط أكثر عمومية حيث لا يتطلب وجود انحناء في كافة الاتجاهات، بل يكفي وجود نمو أدنى للدالة بالنسبة لمجموعة الحلول المثلى ويمكن اعتباره تعميمًا لشرط النمو التربيعي عند استخدام مسافة بريغمان بدلاً من المسافة الإقليدية. ويعطى الشرط وفق:

الدالة f تحقق شرط نمو مسافة بريغمان، أذا وجد $\mu > 0$ حيث

$$f(x) - f(x^*) \geq \mu D_h(x^*, x) \quad (22)$$

لكل $x \in int\ domh$ حيث x^* حل الأمثل للمسألة P . وهو شرط يربط بين فجوة تابع الهدف ومسافة بريغمان نحو الحل الأمثل وهو أضعف من التحدب القوي النسبي حيث يفرض الانحناء فقط بالنسبة لنقطة الحل الأمثل بينما شرط التحدب القوي نسبيا يفرض انحناء معين عند كل نقطتين x, y في كل مكان.

نثبت في المبرهنة الآتية أن شرط نمو بريغمان شرط كافي للتقارب الخطي للخوارزمية متجاوزين الافتراضات القياسية في الأدبيات.

مبرهنة 13:

ليكن الزوج (f, h) ينتمي الى $R_L(C)$ وh تابع بريغمان وf محدبة تماماً ويحقق شرط نمو بريغمان ولتكن $\{x_k\}_{k\in\mathbb{N}}$متتالية معرّفة بالعلاقة(7)، حيث $\lambda \in \left]0, \frac{1}{L}\right[$ وليكن x^* حل أصغري لـf عندئذٍ لكل $k \geq 0$ يكون

$$f(x_{k+1}) - f(x^*) \leq \left[\frac{L}{L+\mu}\right]^k L D_h(x^*, x_0).$$

البرهان.

من العلاقة (11) وحسب تعريف المتتالية $x_{k+1} \in T(x_k)$ فكل$u \in C$نحصل على:

$$f(x_{k+1}) - f(u) \leq \frac{1}{\lambda}[D_h(u, x_k) - D_h(u, x_{k+1})] - \frac{(1 - \lambda L)}{\lambda} D_h(x_{k+1}, x_k). \quad (23)$$

باستبدال u بـ x^* وباختيار $\lambda = \frac{1}{L}$ بـ (23) نحصل على:

$$f(x_{k+1}) - f(x^*) \leq L[D_h(x^*, x_k) - D_h(x^*, x_{k+1})]. \quad (24)$$

من جهة أولى حسب خواص مسافة بريغمان$D_h(.,.) > 0$ فلكل $k \geq 0$ يكون $-D_h(x^*, x_{k+1}) \leq 0$ بالمقارنة المتراجحة (24) نستنتج أنّ:

$$f(x_{k+1}) - f(x^*) \leq LD_h(x^*, x_k). \quad (25)$$

ومن جهة ثانية حسب الشرط A لدينا $f(x_{k+1}) - f(x^*) \leq \mu D_h(x^*, x_{k+1})$ بالمقارنة مع المتراجحة (24) نحصل على:

$$\mu D_h(x^*, x_{k+1}) \leq f(x_k) - f(x^*) \leq L[D_h(x^*, x_k) - D_h(x^*, x_{k+1})].$$

وبإعادة الترتيب يكون لكل $k \in 0$:

$$(\mu + L)D_h(x^*, x_{k+1}) \leq LD_h(x^*, x_k).$$

أي لكل $k \in 0$ يكون

$$D_h(x^*, x_{k+1}) \leq \frac{L}{(L+\mu)} D_h(x^*, x_k).$$

وبتطبيق المتراجحة بشكل متكرر نحصل على

$$D_h(x^*, x_k) \leq \frac{L}{(L+\mu)} D_h(x^*, x_{k-1})$$
$$D_h(x^*, x_{k-1}) \leq \frac{L}{(L+\mu)} D_h(x^*, x_{k-2})$$
$$\vdots$$
$$D_h(x^*, x_1) \leq \frac{L}{(L+\mu)} D_h(x^*, x_0)$$

أي

$$D_h(x^*, x_k) \leq \frac{L}{(L+\mu)} D_h(x^*, x_{k-1}) \leq \left[\frac{L}{(L+\mu)}\right]^2 D_h(x^*, x_{k-2}) \leq \cdots.$$
$$\leq \left[\frac{L}{(L+\mu)}\right]^k D_h(x^*, x_0).$$

ومنه $D_h(x^*, x_k) \leq \left[\frac{L}{(L+\mu)}\right]^k D_h(x^*, x_0)$ (26)

بمقارنة المتراجحتين (25) و (26) نستنتج أنّ:

$$f(x_{k+1}) - f(x^*) \leq \left[\frac{L}{L+\mu}\right]^k LD_h(x^*, x_0).$$

وبذلك يتم المطلوب.■

تثبت المبرهنة 13 التقارب الخطي حيث $0 < \rho = \frac{L}{L+\mu} < 1$ و $C = LD_h(x^*, x_0)$ تحت شروط أضعف أي تحت الشرط نمو بريغمان كبديل أضعف من التحدب القوي النسبي حيث يثبت كفاءة الخوارزمية حتى في غياب الانحناء الصارم لدالة الهدف مما يفتح آفاقا في تدريب الشبكة العصبية العميقة دون أن تتطلب انحناء مثالي عند كل نقطة.

الاستنتاجات والتوصيات:

قدمت المقالة تحليل شامل لخوارزمية انحدار المرآة لحل مسائل الأصغريات المحدبة غير الملساء، حيث توصلت الدراسة إلى أن الخوارزمية تُعد خيارًا فعالًا وأظهرت النتائج تفوقها في تحسين الأداء مقارنة بالطرق التقليدية، حيث أثبتت قدرتها على تحقيق تقارب خطي وعالمي دون الحاجة إلى شروط صارمة. وقد ساهم استخدام نواة بريغمان

ومفهوم السلاسة النسبية في تعزيز مرونة الخوارزمية وكفاءتها، مما يفتح المجال لتطبيقها في مجموعة واسعة من المسائل العددية. ونوصي بإجراء دراسات مستقبلية لتطوير نسخة محسّنة من الخوارزمية بهدف تحسين معدل التقارب وتوسيع نطاق الخوارزمية لتشمل مسائل الأمثليات غير المحدبة.

References:

[1] A. Auslender and M. Teboulle, "Interior gradient and proximal methods for convex and conic optimization," *SIAM Journal on Optimization*, vol. 16, no. 3, pp. 697–725, 2006.
[2] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman divergences," *Journal of Machine Learning Research*, vol. 6, no. Oct, pp. 1705–1749, 2005.
[3] H. H. Bauschke, J. Bolte, and M. Teboulle, "A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications," *Mathematics of Operations Research*, vol. 42, no. 2, pp. 330–348, 2017.
[4] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. New York: Springer, 2017.
[5] A. Beck, A. Ben-Tal, N. Guttmann-Beck, and L. Tetruashvili, "The comirror algorithm for solving nonsmooth constrained convex problems," *Operations Research Letters*, vol. 38, no. 6, pp. 493–498, 2010.
[6] A. Beck and M. Teboulle, "Mirror descent and nonlinear projected subgradient methods for convex optimization," *Operations Research Letters*, vol. 31, no. 3, pp. 167–175, 2003.
[7] A. Ben-Tal, T. Margalit, and A. Nemirovski, "The ordered subsets mirror descent optimization method with applications to tomography," *SIAM Journal on Optimization*, vol. 12, no. 1, pp. 79–108, 2001.
[8] J. M. Bioucas-Dias and M. A. Figueiredo, "Total variation restoration of speckled images using a split-Bregman algorithm," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 3717–3720.
[9] J. Bolte, T. P. Nguyen, J. Peypouquet, and B. W. Suter," From error bounds to the complexity of first-order descent methods for convex functions", * Mathematical Programming *, vol. 165, no. 2, pp. 471‑507, 2017.
[10] J. Bolte, S. Sabach, M. Teboulle, and Y. Vaisbourd, "First order methods beyond convexity and Lipschitz gradient continuity with applications to quadratic inverse problems," *SIAM Journal on Optimization*, vol. 28, no. 3, pp. 2131–2151, 2018.
[11] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 7, no. 3, pp. 200–217, 1967.
[12] M. Burger, "Bregman distances in inverse problems and partial differential equations," in *Advances in Mathematical Modelling, Optimization and Optimal Control*, Cham: Springer, 2016, pp. 3–33.
[13] D. Butnariu, C. Byrne, and Y. Censor, "Redundant axioms in the definition of Bregman functions," *Journal of Convex Analysis*, vol. 10, no. 1, pp. 245–254, 2003.
[14] Y. Censor and A. Lent, "An iterative row-action method for interval convex programming," *Journal of Optimization Theory and Applications*, vol. 34, no. 3, pp. 321–353, 1981.
[15] T. T. Doan, S. Bose, D. H. Nguyen, and C. L. Beck, "Convergence of the iterates in mirror descent methods," *IEEE Control Systems Letters*, vol. 3, no. 1, pp. 114–119, 2019.
[16] W. Krichene, A. Bayen, and P. L. Bartlett, "Accelerated mirror descent in continuous and discrete time," *Advances in Neural Information Processing Systems*, vol. 28, pp. 2845–2853, 2015.

[17] B. Kulis, M. A. Sustik, and I. S. Dhillon, "Low-rank kernel learning with Bregman matrix divergences," *Journal of Machine Learning Research*, vol. 10, no. 2, 2009.
[18] H. Lu, R. M. Freund, and Y. Nesterov, "Relatively smooth convex optimization by first-order methods, and applications," *SIAM Journal on Optimization*, vol. 28, no. 1, pp. 333–354, 2018.
[19] A. Nazin, A. Tremba, and S. Anulova, "Application of the mirror descent method to minimize average losses coming by a Poisson flow," in *Proc. European Control Conf. (ECC)*, 2014, pp. 2194–2197.
[20] A. S. Nemirovsky and D. B. Yudin, *Problem Complexity and Method Efficiency in Optimization*. New York: Wiley, 1983.
[21] A. Nemirovski, "Efficient methods for large-scale convex optimization problems," *Ekonomika i Matematicheskie Metody*, vol. 15, no. 1, 1979.
[22] B. T. Polyak, *Introduction to Optimization*. Optimization Software, Inc., Publications Division, 1987.
[23] R. T. Rockafellar, *Convex Analysis*. Princeton University Press, 1970.
[24] R. T. Rockafellar and R. J. Wets, *Variational Analysis*, 3rd corrected printing. Springer, 2009.
[25] M. Teboulle, "A simplified view of first order methods for optimization," *Mathematical Programming*, vol. 170, no. 1, pp. 67–96, 2018.
[26] Q. Van Nguyen, "Forward-backward splitting with Bregman distances," *Vietnam Journal of Mathematics*, vol. 45, no. 3, pp. 519–539, 2017.
[27] H. Zhang, Y. H. Dai, L. Guo, and W. Peng, "Proximal-like incremental aggregated gradient method with linear convergence under Bregman distance growth conditions," *Mathematics of Operations Research*, vol. 46, no. 1, pp. 61–81, 2021.
[28] Y. Zhou, Y. Liang, and L. Shen, "A simple convergence analysis of Bregman proximal gradient algorithm," *Computational Optimization and Applications*, vol. 73, no. 3, pp. 903–912, 2019.