

VỀ MỘT THUẬT TOÁN TỐI ƯU DỰA TRÊN PHƯƠNG PHÁP
SUBGRADIENT CHO HÀM LIPSCHITZ ĐỊA PHƯƠNG*On an optimality algorithm based on the subgradient method
for locally Lipschitz functions*Phạm Mạnh Chinh⁽¹⁾, Tạ Quang Sơn^{(2)*}⁽¹⁾Trung tâm Giáo dục Nghề nghiệp - Giáo dục Thường xuyên Quận 12⁽²⁾Trường Đại học Sài Gòn**Tóm tắt**

Trong bài báo này, chúng tôi quan tâm đến phương pháp subgradient thông qua một dãy lặp có sử dụng dưới vi phân Clarke. Phương pháp này được dùng để xây dựng thuật toán tìm nghiệm tối ưu của bài toán tối ưu không lồi với hàm mục tiêu là hàm Lipschitz địa phương. Các kết quả đạt được về sự hội tụ, độ chính xác và điều kiện dừng của thuật toán sẽ được giới thiệu. Phương pháp sử dụng trong nghiên cứu này mở rộng các kết quả được phát triển bởi S. Boyd cho bài toán tối ưu lồi được giới thiệu trong [1].

Từ khóa: Phương pháp subgradient; Dưới vi phân Clarke; Sự hội tụ của dãy lặp; Điều kiện dừng.

Abstract

In this paper, we are concerned with a subgradient method based on an iterative sequence involving the Clarke subdifferential. This method is used to construct an algorithm for finding optimal solutions of a nonconvex optimization problem whose objective function is locally Lipschitz. The obtained results regarding the convergence, accuracy, and stopping conditions of the algorithm are presented. The method used in this study extends the framework developed by S. Boyd for convex optimization problems introduced in [1].

Keywords: Subgradient method; Clarke subdifferential; Iterative convergence; Stopping conditions.

1. Giới thiệu

Xét bài toán tối ưu không ràng buộc

$$\begin{aligned} (P) \text{ Min } & f(x) \\ \text{s. t. } & x \in \mathbb{R}^n, \end{aligned}$$

trong đó $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Đối với bài toán lồi khả vi, phương pháp gradient tìm nghiệm tối ưu được áp dụng với bước lặp có dạng

$$x^{(k+1)} = x^{(k)} - \alpha_k \nabla f(x^{(k)}), \quad (1.1)$$

với $-\nabla f(x^{(k)})$ là hướng giảm sâu nhất, $\alpha_k > 0$ là bước nhảy của thuật toán. Trong trường hợp bài toán lồi nhưng không khả vi, bước lặp của thuật toán được cho bởi

$$x^{(k+1)} = x^{(k)} - \alpha_k u^{(k)}. \quad (1.2)$$

*Tác giả liên hệ: taquangson@sgu.edu.vn

Trong đó, $x^{(k)}$ là phần tử thứ k , $u^{(k)} \in \partial f(x^{(k)})$ là một vec-tơ subgradient nào đó của hàm f tại $x^{(k)}$ và $\alpha_k > 0$ là bước nhảy thứ k của thuật toán.

Tuy nhiên, đối với phương pháp subgradient, $-u^{(k)}$ có thể không phải hướng giảm của hàm f tại $x^{(k)}$. Hơn nữa, dù $-u^{(k)}$ là hướng giảm tại $x^{(k)}$ thì bước nhảy α_k vẫn có thể khiến $f(x^{(k+1)}) > f(x^{(k)})$ (xem [1]). Vì vậy phương pháp subgradient nói chung chưa bảo đảm làm cho các giá trị của bài toán giảm dần. Do đó, cần phải thực hiện quy trình tìm điểm tốt nhất tại mỗi bước lặp, tức là tìm điểm làm giá trị của hàm mục tiêu nhỏ nhất tại bước thứ k . Cụ thể, ở bước lặp thứ k , ta chọn giá trị nhỏ nhất của hai giá trị $f_{\text{best}}^{(k-1)}$ và $f(x^{(k)})$ và ta đặt $i_{\text{best}}^{(k)} = k$ nếu $f(x^{(k)}) = f_{\text{best}}^{(k)}$ trong đó

$$f_{\text{best}}^{(k)} = \min \{ f(x^{(1)}), \dots, f(x^{(k)}) \}.$$

là giá trị tốt nhất thu được của thuật toán ở bước lặp thứ k . Nếu giả thiết f bị chặn dưới thì khi $f_{\text{best}}^{(k)}$ giảm dần với $k \rightarrow \infty$, ta có $\lim_{k \rightarrow \infty} f_{\text{best}}^{(k)}$ tồn tại. Các kết quả này đã được giới thiệu rõ trong [1].

Trong bài báo này, bài toán (P) có f là hàm Lipschitz không khả vi. Đối chiếu các dãy lặp (1.1), (1.2) với chú ý rằng dưới vi phân Clarke tại $x^{(k)}$ là bao lồi của các $\lim_{k \rightarrow \infty} \nabla f(x_n^{(k)})$, trong đó $x_n^{(k)} \rightarrow x^{(k)}$ với $x_n^{(k)} \in D$ là tập các điểm mà hàm f khả vi, chúng tôi đề xuất dùng dãy lặp (1.2) với $u^{(k)} \in \partial^c f(x^{(k)})$.

Các cách chọn bước nhảy α_k dưới đây được tham khảo từ [1], gồm có:

- *Bước nhảy hằng số*: $\alpha_k = \alpha$ với α là một hằng số dương.
- *Bước nhảy độ dài hằng số*: $\alpha_k = \frac{\gamma}{\|u^{(k)}\|}$, trong đó $\gamma > 0$. Với cách chọn bước nhảy này, khoảng cách giữa điểm thứ $k+1$ và điểm thứ k xác định bởi phương pháp subgradient là hằng số. Tức là

$$\|x^{(k+1)} - x^{(k)}\| = \alpha_k \|u^{(k)}\| = \gamma.$$

- *Bước nhảy khả tổng bình phương nhưng không khả tổng*: Bước nhảy α_k thoả các tính chất sau

$$\alpha_k \geq 0, \quad \sum_{k=1}^{\infty} \alpha_k^2 < \infty, \quad \sum_{k=1}^{\infty} \alpha_k = \infty.$$

Phương pháp subgradient tìm nghiệm của bài toán tối ưu đã được nhiều tác giả quan tâm trong nhiều năm qua. Phương pháp này được Shor và các cộng sự đề xuất từ những năm

1960 và đã được giới thiệu trong nhiều tài liệu với những phát triển [2], [3], [4], [5]. Liên quan đến tốc độ hội tụ của thuật toán, gần đây nhiều tác giả đã quan tâm cải tiến phương pháp này [6], [7]. Mục đích của chúng tôi là mở rộng các kết quả về phương pháp subgradient được Stephen Boyd giới thiệu trong tài liệu [1] dành cho các bài toán tối ưu lồi. Từ các kết quả liên quan đến dưới vi phân của hàm lồi, chúng tôi mở rộng thuật toán bằng cách dùng dưới vi phân cho hàm Lipschitz địa phương để tìm nghiệm tối ưu cho bài toán không lồi.

Bài báo được tổ chức như sau: Trong mục 2, chúng tôi trình bày một số kiến thức cơ bản liên quan đến Giải tích Lipschitz và dưới vi phân Clarke cùng với một số tính chất liên quan. Trong mục 3, chúng tôi khảo sát sự hội tụ của thuật toán với các cách chọn bước nhảy khác nhau. Điều kiện dừng của thuật toán cũng được nghiên cứu. Một ví dụ minh họa cũng được giới thiệu.

2. Kiến thức chuẩn bị

Trong bài báo này, các kết quả được xét trên không gian $\mathbb{R}^n$ với tích trong $\langle \cdot, \cdot \rangle$ và chuẩn

$\| \cdot \|$ định nghĩa bởi $\|x\| = \langle x, x \rangle^{\frac{1}{2}} = \sqrt{\sum_{i=1}^n x_i^2}$. Với $x \in \mathbb{R}^n$, $U(x)$ được gọi là một lân cận của x nếu tồn tại quả cầu mở $B(x, r) = \{y \in \mathbb{R}^n \mid \|y - x\| < r\} \subset U(x)$. Bao lồi của tập A , kí hiệu là $\text{co}(A)$, cho bởi

$$\text{co}(A) := \{x \in \mathbb{R}^n \mid x = \sum_{i=1}^k \lambda_i x_i, \sum_{i=1}^k \lambda_i = 1, x_i \in A, \lambda_i \geq 0, k \in \mathbb{N}\}. \quad (2.1)$$

Hàm số f xác định trên $\mathbb{R}^n$ gọi là hàm Lipschitz địa phương tại điểm $x \in \mathbb{R}^n$ nếu tồn tại lân cận $U(x) \subset \mathbb{R}^n$ và hằng số K sao cho

$$|f(y) - f(z)| \leq K \|y - z\|, \forall y, z \in U(x).$$

Ta nói f là hàm Lipschitz địa phương trên $\mathbb{R}^n$ nếu f là hàm Lipschitz địa phương tại mọi điểm $x \in \mathbb{R}^n$.

Cho hàm $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ và vec-tơ $d \in \mathbb{R}^n \setminus \{0\}$. Xét $x_0 \in \text{dom} f$ sao cho $x_0 + \lambda d \in \text{dom} f$ với $\lambda > 0$. Đạo hàm theo hướng d tại x_0 của hàm f được định nghĩa và kí hiệu bởi

$$f'(x_0; d) := \lim_{\lambda \rightarrow 0^+} \frac{f(x_0 + \lambda d) - f(x_0)}{\lambda}.$$

Một vec-tơ $d \in \mathbb{R}^n \setminus \{0\}$ gọi là hướng giảm của hàm f tại $x \in \text{dom} f$ nếu đạo hàm theo hướng $f'(x; d)$ tồn tại và $f'(x; d) < 0$. Ta cũng có tính chất sau: Cho hàm $f: \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ và $x \in \text{dom} f$. Giả sử $d \in \mathbb{R}^n \setminus \{0\}$ là hướng giảm của hàm f tại x . Khi đó tồn tại $\varepsilon > 0$ sao cho $x + td \in \text{dom} f$ và

$$f(x + td) < f(x), \forall t \in (0, \varepsilon].$$

Trong trường hợp f là hàm Lipschitz địa phương, khái niệm đạo hàm theo hướng được mở rộng thành khái niệm đạo hàm Clarke. Các kết quả dưới đây có thể được tìm thấy trong tài liệu [8].

Đạo hàm Clarke tại x_0 theo hướng d được định nghĩa bởi

$$f^c(x_0; d) := \limsup_{x \rightarrow x_0, t \downarrow 0} \frac{f(x+td) - f(x)}{t}.$$

Hàm $f: \mathbb{R}^n \rightarrow \mathbb{R}$ được gọi là chính quy tại x nếu với mọi $d \in \mathbb{R}^n$, đạo hàm theo hướng $f'(x; d)$ tồn tại và $f^c(x; d) = f'(x; d)$, với mọi $d \in \mathbb{R}^n$.

Từ đây, dưới vi phân Clarke cho hàm Lipschitz địa phương được kí hiệu và định nghĩa bởi

$$\partial^c f(x_0) := \{u \in \mathbb{R}^n \mid f^c(x_0; d) \geq \langle u, d \rangle, \forall d \in \mathbb{R}^n\}. \quad (2.2)$$

Mỗi vec-tơ u trong công thức (2.2) gọi là vec-tơ dưới đạo hàm Clarke của hàm f tại x_0 . Ta cũng có

$$f^c(x_0; d) = \max_{u \in \partial^c f(x_0)} \langle u, d \rangle.$$

Chú ý rằng dưới vi phân Clarke còn được biểu diễn bởi

$$\partial^c f(x_0) = \text{co}\{\xi \in \mathbb{R}^n \mid \xi = \lim_{k \rightarrow \infty} \nabla f(x_k), x_k \rightarrow x_0, x_k \in D\},$$

trong đó D là tập các điểm mà tại đó f khả vi [8].

Nếu hàm f là hàm lồi trên lân cận U của x thì $\partial^c f(x) = \partial f(x)$ và $f^c(x; d) = f'(x; d)$ với mọi $d \in \mathbb{R}^n$. Hơn nữa, nếu $x \in \text{int}(\text{dom}(f))$, f là hàm khả vi trên U thì $\partial^c f(x) = \{\nabla f(x)\}$.

Mệnh đề 2.1 ([8], Mệnh đề 2.1.2, trang 27) Cho f là hàm Lipschitz địa phương tại $x \in \mathbb{R}^n$ với K là hằng số Lipschitz, khi đó $\|\xi\| \leq K$ với mọi $\xi \in \partial^c f(x)$.

Với bài toán

$$(P) \begin{array}{ll} \text{Min} & f(x) \\ \text{s. t.} & x \in \mathbb{R}^n, \end{array}$$

trong đó $f: \mathbb{R}^n \rightarrow \mathbb{R}$ là hàm Lipschitz địa phương. Nếu f có một cực tiểu địa phương tại x thì $0 \in \partial^c f(x)$.

Một số định lí sau đây nói về các phép toán dưới vi phân Clarke. Cho $f: \mathbb{R}^n \rightarrow \mathbb{R}$ là hàm Lipschitz địa phương tại $x \in \mathbb{R}^n$ và $\alpha \in \mathbb{R}$, khi đó

$$\partial^c(\alpha f)(x) = \alpha \partial^c f(x).$$

Với $f_i: \mathbb{R}^n \rightarrow \mathbb{R}, i = \overline{1, n}$ là các hàm Lipschitz địa phương tại x xác định trên $\mathbb{R}^n$. Khi đó

$$\partial^c \left(\sum_{i=1}^n f_i \right) (x) \subset \sum_{i=1}^n \partial^c f_i(x).$$

Nếu mỗi hàm f_i chính quy tại x thì

$$\partial^c \left(\sum_{i=1}^n f_i \right) (x) = \sum_{i=1}^n \partial^c f_i(x).$$

Định lí 2.1 ([8], Định lí 2.3.7, trang 41) Cho hàm $f: \mathbb{R}^n \rightarrow \mathbb{R}$ là hàm Lipschitz địa phương trên tập mở chứa đoạn $[x, y]$ với $x \in \mathbb{R}^n, y \in U(x)$. Khi đó tồn tại $u \in (x, y)$ sao cho

$$f(y) - f(x) \in \langle \partial^c f(u), y - x \rangle.$$

3. Các kết quả chính

3.1. Sự hội tụ của thuật toán

Dựa trên các kết quả đạt được về chặn trên của $f_{\text{best}}^{(k)} - f^*$ được giới thiệu trong [1] và [10], chúng tôi đề xuất bổ đề sau về chặn trên đó trong việc giải bài toán tối ưu không lồi (P) dùng dãy lặp (1.2).

$$x^{(k+1)} = x^{(k)} - \alpha_k u^{(k)}.$$

Bổ đề 3.1 Xét bài toán (P) với $f: \mathbb{R}^n \rightarrow \mathbb{R}$ là hàm Lipschitz địa phương. Giả sử tồn tại x^* là nghiệm tối ưu của bài toán, đặt $f(x^*) = f^*$. Giả sử thêm rằng tồn tại số thực dương R thỏa

$$\|x^{(1)} - x^*\| \leq R.$$

Với $K(i)$ là hằng số Lipschitz của hàm mục tiêu tại bước thứ i , ta đặt $G = \max\{K(i), i = \overline{1, k}\}$. Khi đó với $\varepsilon > 0$ tùy ý, ta có

$$f_{\text{best}}^{(k)} - f^* \leq \frac{R^2 + \sum_{i=1}^k \alpha_i^2 G^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon.$$

Chứng minh.

Tại bước nhảy thứ k , đặt $\bar{x}_n^{(k)} = \frac{1}{n} x^* + \left(1 - \frac{1}{n}\right) x^{(k)}$ thì $\bar{x}_n^{(k)} \xrightarrow{n \rightarrow \infty} x^{(k)}$. Giả sử có dãy $\{u_n^{(k)}\}_{n \in \mathbb{N}}$ thỏa $u_n^{(k)} \in \partial^c f(\bar{x}_n^{(k)})$, khi đó tồn tại dãy con $\{u_{n_m}^{(k)}\}$ của $\{u_n^{(k)}\}$ sao cho $u_{n_m}^{(k)}$ hội tụ về $u^{(k)} \in \partial^c f(x^{(k)})$.

Với f là hàm Lipschitz địa phương, áp dụng Định lý 2.1, ta có

$$f^* - f(x^{(k)}) = \langle u_n^{(k)}, x^* - x^{(k)} \rangle.$$

Mà $\langle u_n^{(k)}, x^* - x^{(k)} \rangle \xrightarrow{n \rightarrow \infty} \langle u^{(k)}, x^* - x^{(k)} \rangle$ nên với mọi $\varepsilon > 0$, tồn tại $N \in \mathbb{N}$ sao cho

$$\left| \langle u_n^{(k)}, x^* - x^{(k)} \rangle - \langle u^{(k)}, x^* - x^{(k)} \rangle \right| < \varepsilon, \forall n > N.$$

Hay

$$-\varepsilon < \langle u_n^{(k)}, x^* - x^{(k)} \rangle - \langle u^{(k)}, x^* - x^{(k)} \rangle < \varepsilon.$$

Mà $f^* - f(x^{(k)}) = \langle u_n^{(k)}, x^* - x^{(k)} \rangle$ nên

$$f^* - f(x^{(k)}) + \varepsilon \geq \langle u^{(k)}, x^* - x^{(k)} \rangle, u^{(k)} \in \partial^c f(x^{(k)}). \quad (3.1)$$

Khi đó

$$\begin{aligned} & \|x^{(k+1)} - x^*\|^2 \\ &= \|x^{(k)} - \alpha_k u^{(k)} - x^*\|^2 \\ &= \|x^{(k)} - x^*\|^2 - 2\alpha_k \langle u^{(k)}, x^{(k)} - x^* \rangle + \alpha_k^2 \|u^{(k)}\|^2 \\ &\leq \|x^{(k)} - x^*\|^2 - 2\alpha_k (f(x^{(k)}) - f^* - \varepsilon) + \alpha_k^2 \|u^{(k)}\|^2 \\ &\leq \|x^{(k-1)} - x^*\|^2 - 2\alpha_k (f(x^{(k)}) - f^* - \varepsilon) - 2\alpha_{k-1} (f(x^{(k-1)}) - f^* - \varepsilon) + \alpha_k^2 \|u^{(k)}\|^2 + \alpha_{k-1}^2 \|u^{(k-1)}\|^2 \\ &\leq \dots \\ &\leq \|x^{(1)} - x^*\|^2 - 2 \sum_{i=1}^k \alpha_i (f(x^{(i)}) - f^* - \varepsilon) + \sum_{i=1}^k \alpha_i^2 \|u^{(i)}\|^2. \end{aligned}$$

Với giả thiết $\|x^{(1)} - x^*\| \leq R$ và $\|x^{(k+1)} - x^*\|^2 \geq 0$, ta có

$$0 \leq R^2 - 2 \sum_{i=1}^k \alpha_i (f(x^{(i)}) - f^* - \varepsilon) + \sum_{i=1}^k \alpha_i^2 \|u^{(i)}\|^2,$$

tương đương

$$2 \sum_{i=1}^k \alpha_i (f(x^{(i)}) - f^* - \varepsilon) \leq R^2 + \sum_{i=1}^k \alpha_i^2 \|u^{(i)}\|^2 \quad (3.2).$$

Mặt khác, ta có

$$\sum_{i=1}^k \alpha_i (f(x^{(i)}) - f^* - \varepsilon) \geq \min_{i=1, k} \{f(x^{(i)}) - f^* - \varepsilon\} \sum_{i=1}^k \alpha_i.$$

Mà

$$\begin{aligned} \min_{i=1,\overline{k}} \left\{ f(x^{(i)}) - f^* - \varepsilon \right\} \sum_{i=1}^k \alpha_i &= \left(\min_{i=1,\overline{k}} f(x^{(i)}) - f^* - \varepsilon \right) \sum_{i=1}^k \alpha_i \\ &= \left(f_{\text{best}}^{(k)} - f^* - \varepsilon \right) \sum_{i=1}^k \alpha_i. \end{aligned}$$

Vậy nên

$$2 \sum_{i=1}^k \alpha_i \left(f_{\text{best}}^{(k)} - f^* - \varepsilon \right) \leq R^2 + \sum_{i=1}^k \alpha_i^2 \|u^{(i)}\|^2.$$

Suy ra

$$f_{\text{best}}^{(k)} - f^* \leq \frac{R^2 + \sum_{i=1}^k \alpha_i^2 \|u^{(i)}\|^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon. \quad (3.3)$$

Vì $u^{(i)} \in \partial^c f(x^{(i)})$ nên theo Mệnh đề 2.1 ta có $\|u^{(i)}\| \leq K(i)$ với $K(i)$ là hằng số Lipschitz của hàm mục tiêu tại bước thứ i . Chọn $G = \max\{K(i), i=1,\overline{k}\}$, ta có $\|u^{(i)}\| \leq G, \forall i=1,\overline{k}$. Khi đó (3.3) trở thành

$$f_{\text{best}}^{(k)} - f^* \leq \frac{R^2 + \sum_{i=1}^k \alpha_i^2 G^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon. \quad (3.4) \blacksquare$$

Nếu f là hàm lồi và giả sử tồn tại số G sao cho

$$\|u^{(k)}\| \leq G, u^{(k)} \in \partial f(x^{(k)})$$

với mọi k thì với ε đủ bé, ta có kết quả được đề cập trong [1]

$$f_{\text{best}}^{(k)} - f^* \leq \frac{R^2 + G^2 \sum_{i=1}^k \alpha_i^2}{2 \sum_{i=1}^k \alpha_i}.$$

Bằng cách chọn bước nhảy hằng số hoặc bước nhảy độ dài hằng số, giá trị của $f_{\text{best}}^{(k)} - f^*$ sẽ hội tụ về một lượng tối ưu nào đó. Còn nếu ta chọn bước nhảy khả tổng bình phương nhưng không khả tổng thì $f_{\text{best}}^{(k)} \rightarrow f^*$ với k đủ lớn. Từ đó, ta có cơ sở để dùng phương pháp subgradient với dãy lặp (1.2) tìm nghiệm tối ưu của bài toán (P). Cụ thể, ta có các định lí sau:

Định lý 3.1 Xét bài toán (P) trong đó $f: \mathbb{R}^n \rightarrow \mathbb{R}$ là hàm Lipschitz địa phương với các giả thiết của Bổ đề 3.1. Với bước nhảy $\alpha_k = \alpha$ trong đó $\alpha > 0$ là hằng số cho trước, khi đó

$$f_{\text{best}}^{(k)} - f^* \leq \frac{G^2 \alpha}{2},$$

với k đủ lớn. Hơn nữa

$$f_{\text{best}}^{(k)} - f^* \leq \alpha G^2$$

sau ít nhất n bước nào đó.

Chúng minh.

Từ (3.4), với $\alpha_k = \alpha > 0$, ta có

$$\begin{aligned} f_{\text{best}}^{(k)} - f^* &\leq \frac{R^2 + \sum_{i=1}^k \alpha_i^2 G^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon \\ &= \frac{R^2 + k \alpha^2 G^2}{2k\alpha} + \varepsilon. \end{aligned}$$

Khi k đủ lớn và ε đủ bé, ta có

$$f_{\text{best}}^{(k)} - f^* \leq \frac{R^2}{2k\alpha} + \frac{\alpha G^2}{2} + \varepsilon \rightarrow \frac{\alpha G^2}{2}.$$

Hơn nữa, ta có

$$\frac{R^2}{2k\alpha} \leq \frac{\alpha G^2}{2} \Leftrightarrow k \geq \frac{R^2}{\alpha^2 G^2}.$$

Khi đó với ε đủ bé thì

$$f_{\text{best}}^{(k)} - f^* \leq \alpha G^2, \forall k \geq \frac{R^2}{\alpha^2 G^2}.$$

Từ đó, ta thấy bằng cách chọn α phù hợp, $f_{\text{best}}^{(k)}$ có thể hội tụ về f^* sau ít nhất n bước nhảy với $n = \frac{R^2}{\alpha^2 G^2} \in \mathbb{N}$. ■

Định lý 3.2 Xét bài toán (P) trong đó $f: \mathbb{R}^n \rightarrow \mathbb{R}$ là hàm Lipschitz địa phương với các giả thiết của Bổ đề 3.1. Với bước nhảy $\alpha_k = \frac{\gamma}{\|u^{(k)}\|}$ trong đó $\gamma > 0$ là hằng số cho trước,

khi đó $f_{\text{best}}^{(k)} - f^* \leq \frac{G\gamma}{2}$, khi k đủ lớn.

Chứng minh.

Từ (3.3), ta có

$$\begin{aligned} f_{\text{best}}^{(k)} - f^* &\leq \frac{R^2 + \sum_{i=1}^k \alpha_i^2 \|u^{(i)}\|^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon \\ &= \frac{R^2 + \sum_{i=1}^k \gamma^2}{2 \sum_{i=1}^k \frac{\gamma}{\|u^{(i)}\|}} + \varepsilon. \end{aligned}$$

Mà $\|u^{(i)}\| \leq G$ với mọi i , do đó

$$f_{\text{best}}^{(k)} - f^* \leq \frac{R^2 + k\gamma^2}{2\gamma k/G} + \varepsilon.$$

Với ε đủ bé và k đủ lớn thì

$$\frac{R^2 + k\gamma^2}{2\gamma k/G} + \varepsilon \rightarrow \frac{G\gamma}{2}.$$

Khi đó $f_{\text{best}}^{(k)} - f^* \leq \frac{G\gamma}{2}$. ■

Định lý 3.3 Xét bài toán (P) trong đó $f: \mathbb{R}^n \rightarrow \mathbb{R}$ là hàm Lipschitz địa phương với các giả thiết của Bổ đề 3.1. Xét bước nhảy $\alpha_k \geq 0$ thoả

$$\|\alpha\|^2 = \sum_{k=1}^{\infty} \alpha_k^2 < \infty, \quad \sum_{k=1}^{\infty} \alpha_k = \infty.$$

Khi đó, $f_{\text{best}}^{(k)}$ hội tụ về f^* khi k đủ lớn và với $\varepsilon > 0$ đủ bé.

Chứng minh.

Từ (3.4), ta có

$$\begin{aligned} f_{\text{best}}^{(k)} - f^* &\leq \frac{R^2 + G^2 \sum_{i=1}^k \alpha_i^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon \\ &\leq \frac{R^2 + G^2 \|\alpha\|^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon. \end{aligned}$$

Khi k đủ lớn, ta có $\sum_{i=1}^k \alpha_i^2 \rightarrow \|\alpha\|^2$ là hằng số và $2\sum_{i=1}^k \alpha_i \rightarrow \infty$ nên với ε đủ bé thì

$$f_{\text{best}}^{(k)} - f^* \rightarrow 0 \quad \text{hay} \quad f_{\text{best}}^{(k)} \rightarrow f^* . \blacksquare$$

3.2. Độ chính xác và điều kiện dừng

Ở phần trước, từ Bổ đề 3.1, ta đã chỉ ra được một chặn trên của $f_{\text{best}}^{(k)} - f^*$ là

$$f_{\text{best}}^{(k)} - f^* \leq \frac{R^2 + G^2 \sum_{i=1}^k \alpha_i^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon.$$

Trong mục này, ta quan tâm đến cách chọn dãy các bước nhảy và số bước nhảy để chặn trên đó là nhỏ nhất.

Trong [1], [10] và [11], các tác giả đã tìm nghiệm tối ưu của bài toán (P) với hàm mục tiêu f là hàm lồi bằng phương pháp phép chiếu subgradient. Bằng cách định nghĩa ε -nghiệm của bài toán là vec-tơ $x \in \mathbb{R}^n$ thỏa $f(x) - f^* \leq \varepsilon$, các tác giả đã ước lượng được số bước nhảy để thuật toán đạt được ε -nghiệm. Nói cách khác, với $\varepsilon > 0$ là độ chính xác của thuật toán, các tác giả đã đưa ra những nhận xét về số bước nhảy k sao cho $f_{\text{best}}^{(k)} - f^* \leq \varepsilon$.

Dựa trên những kết quả đó, cho trước số dương η là độ chính xác của thuật toán. Nếu ta coi biểu thức (3.4) là điều kiện dừng của thuật toán thì câu hỏi đặt ra là sau ít nhất bao nhiêu bước lặp, $f_{\text{best}}^{(k)}$ sẽ hội tụ về f^* với độ chính xác η ? Để trả lời câu hỏi này, chúng tôi đưa ra mệnh đề sau:

Mệnh đề 3.1 Xét bài toán (P), áp dụng phương pháp subgradient với dãy lặp (1.2) và các giả thiết ở Bổ đề 3.1. Gọi N là số bước nhảy để thuật toán có độ chính xác η , tức là $f(x^{(N)}) - f^* \leq \eta$. Khi đó

$$N \geq \frac{R^2 G^2}{(\eta - \varepsilon)^2}.$$

Chứng minh.

Từ biểu thức (3.4), ta có một chặn trên của $f_{\text{best}}^{(k)} - f^*$ là

$$\frac{R^2 + \sum_{i=1}^k \alpha_i^2 G^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon.$$

Ta sẽ cực tiểu hoá biểu thức này như sau: Xét hàm $\varphi: \mathbb{R}_{++}^k \rightarrow \mathbb{R}$ cho bởi

$$\varphi(\alpha) = \frac{R^2 + \sum_{i=1}^k \alpha_i^2 G^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon.$$

Ta thấy $\varphi(\alpha)$ là hàm khả vi, do đó $\nabla \varphi(\alpha) = \left[\frac{\partial(\alpha)}{\partial \alpha_i} \right]_{i=1, \overline{k}}$, trong đó

$$\frac{\partial(\alpha)}{\partial \alpha_i} = \frac{2\alpha_i G^2 \sum_{i=1}^k \alpha_i - \left(R^2 + G^2 \sum_{i=1}^k \alpha_i^2 \right)}{2 \left(\sum_{i=1}^k \alpha_i \right)^2}, i = \overline{1, k}.$$

Cho $\nabla \varphi(\alpha) = 0$, suy ra

$$\alpha_i = \frac{R^2 + G^2 \sum_{i=1}^k \alpha_i^2}{2G^2 \sum_{i=1}^k \alpha_i}$$

với mọi $i \in \{1; 2; \dots; k\}$. Theo nguyên lí Fermat thì $\varphi(\alpha)$ đạt cực tiểu khi

$$\alpha_1 = \alpha_2 = \dots = \alpha_k = \frac{R^2 + G^2 \sum_{i=1}^k \alpha_i^2}{2G^2 \sum_{i=1}^k \alpha_i}.$$

Để đơn giản, ta đặt $\alpha_i = \alpha$ với mọi $i = \overline{1, k}$. Khi đó từ (3.4), ta có

$$f_{\text{best}}^{(k)} - f^* \leq \frac{R^2 + G^2 k \alpha^2}{2k\alpha} + \varepsilon.$$

Xét hàm $\phi: \mathbb{R} \rightarrow \mathbb{R}$ cho bởi $\phi(\alpha) = \frac{R^2 + G^2 k \alpha^2}{2k\alpha} + \varepsilon$, ta có

$$\phi'(\alpha) = \frac{-R^2}{2k\alpha^2} + \frac{G^2}{2}.$$

Cho $\phi'(\alpha) = 0$, suy ra $\alpha = \frac{R}{G\sqrt{k}}$. Theo nguyên lí Fermat thì $\phi(\alpha)$ đạt cực tiểu khi

$$\alpha = \frac{R}{G\sqrt{k}}. \text{ Khi đó}$$

$$\phi_{\min}(\alpha) = \frac{RG}{\sqrt{k}} + \varepsilon.$$

Với N là số bước nhảy để thuật toán có độ chính xác η , ta có

$$\frac{RG}{\sqrt{N}} + \varepsilon \leq \eta \quad \text{hay} \quad N \geq \frac{R^2 G^2}{(\eta - \varepsilon)^2}. \quad \blacksquare$$

Vậy nếu ta chọn biểu thức (3.4) làm điều kiện dừng của thuật toán, thì số bước nhảy ít nhất có thể để thuật toán có độ chính xác η là $\frac{R^2 G^2}{(\eta - \varepsilon)^2}$, dù cho biểu thức của bước nhảy α_k ta chọn là gì đi chăng nữa. Hơn nữa, nếu ta muốn số bước nhảy của thuật toán là nhỏ nhất $N = \frac{R^2 G^2}{(\eta - \varepsilon)^2} \in \mathbb{N}$ thì có thể chọn bước nhảy của dãy lặp (1.2) có dạng là

$$\alpha_i = \frac{R}{G\sqrt{N}} = \frac{\eta - \varepsilon}{G^2}, \quad \forall i = \overline{1, k}.$$

Việc chọn bước nhảy $\alpha_i = \frac{\eta - \varepsilon}{G^2}$ với mọi $i = \overline{1, k}$ giúp dãy lặp của thuật toán không bị phụ thuộc vào số bước nhảy N và chặn trên R của khoảng cách giữa điểm đầu và nghiệm tối ưu của thuật toán. Đây là những giá trị thường ta không biết trước, trong khi đó chặn trên G của $\|u^{(i)}\|$ lại ước lượng được trong quá trình thực hiện thuật toán.

Ngoài ra, từ (3.2), một chặn dưới của f^* cũng có thể được dùng làm điều kiện dừng của thuật toán đang xét. Chặn dưới đó được xác định như sau

$$\begin{aligned} 2 \sum_{i=1}^k \alpha_i (f(x^{(i)}) - f^* - \varepsilon) &\leq R^2 + \sum_{i=1}^k \alpha_i^2 \|u^{(i)}\|^2 \\ \Leftrightarrow -2 \sum_{i=1}^k \alpha_i (f^* + \varepsilon) &\leq R^2 + \sum_{i=1}^k \alpha_i^2 \|u^{(i)}\|^2 - 2 \sum_{i=1}^k \alpha_i f(x^{(i)}) \\ \Leftrightarrow f^* &\geq l_k, \end{aligned}$$

với

$$l_k := \frac{2 \sum_{i=1}^k \alpha_i f(x^{(i)}) - R^2 - \sum_{i=1}^k \alpha_i^2 \|u^{(i)}\|^2}{2 \sum_{i=1}^k \alpha_i} - \varepsilon.$$

Từ đây, ta xác định được một chặn dưới l_k của f^* tại bước thứ k . Dãy các chặn dưới $\{l_k\}$ này không nhất thiết là dãy tăng, do đó chặn dưới tốt nhất, kí hiệu là $l_{\text{best}}^{(k)}$, có thể được chọn như sau

$$l_{\text{best}}^{(k)} = \max\{l_1, l_2, \dots, l_k\}.$$

Khi đó, điều kiện để dừng thuật toán là $f_{\text{best}}^{(k)} - l_{\text{best}}^{(k)}$ nhỏ hơn một ngưỡng nào đó. Ta thấy chặn dưới l_k này tốt hơn chặn dưới được suy ra từ (3.4)

$$f_{\text{best}}^{(k)} - f^* \leq \frac{R^2 + \sum_{i=1}^k \alpha_i^2 G^2}{2 \sum_{i=1}^k \alpha_i} + \varepsilon$$

$$\Leftrightarrow f^* \geq \frac{2 \sum_{i=1}^k \alpha_i f_{\text{best}}^{(k)} - R^2 - \sum_{i=1}^k \alpha_i^2 G^2}{2 \sum_{i=1}^k \alpha_i} - \varepsilon,$$

nhưng tốc độ hội tụ của thuật toán vẫn rất chậm. Do đó phương pháp subgradient thường không có điều kiện dừng cố định nào.

Chúng tôi minh hoạ phương pháp subgradient được đề cập trong bài báo này thông qua một ví dụ số. Trước hết, các bước của phương pháp subgradient được thực hiện như trong Bảng 1.

Bảng 1: Thuật toán subgradient với bước nhảy hằng số

Đầu vào:	Bước nhảy $\alpha_k = \alpha$, số bước nhảy N .
Khởi tạo:	Chọn điểm đầu $x^{(1)} \in \mathbb{R}^n$.
Các bước chính:	Với mọi $0 < k \leq N$, ta thực hiện các bước sau:
	1. Tính $u^{(k)}$. Nếu $u^{(k)} = 0$ thì $x^* = x^{(k)}$, ta dừng thuật toán;
	2. Tìm $x^{(k+1)} = x^{(k)} - \alpha_k u^{(k)}$, đặt $k := k+1$ rồi trở lại bước 1.

Thuật toán này áp dụng vào ví dụ minh hoạ như sau:

Ví dụ 3.1 Tìm nghiệm tối ưu của bài toán

$$(P_1) \text{Min } f(x) = \begin{cases} x+2, & x > 2 \\ x^2, & |x| \leq 2 \\ 2-x, & x < -2 \end{cases}$$

$$\text{s.t. } x \in \mathbb{R},$$

Đặt $f_1(x) = x+2$, $f_2(x) = x^2$, $f_3(x) = 2-x$.

Khi đó trên $\mathbb{R}$, ta có $\nabla f_1(x) = 1$, $\nabla f_2(x) = 2x$, $\nabla f_3(x) = -1$.

Với $x \in (-\infty; -2) \cup (-2; 2) \cup (2; +\infty)$, hàm f khả vi nên $\partial^c f(x) = \{\nabla f(x)\}$.

Tại $x = 2$, ta có

$$\partial^c f(2) = \text{co}\{\xi \in \mathbb{R} \mid \xi = \lim_{k \rightarrow \infty} \nabla f(x_k), x_k \rightarrow 2, x_k \in D\} = [1; 4],$$

với D là tập các điểm khả vi của f .

Tương tự, tại $x = -2$, ta có

$$\partial^c f(-2) = \text{co}\{\xi \in \mathbb{R} \mid \xi = \lim_{k \rightarrow \infty} \nabla f(x_k), x_k \rightarrow -2, x_k \in D\} = [-4; -1].$$

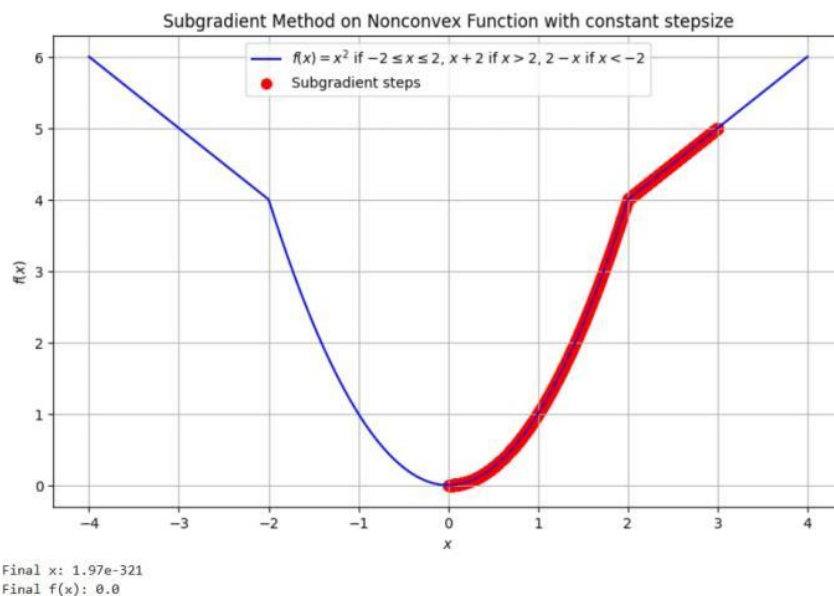
Tóm lại, ta có

$$\partial^c f(x) = \begin{cases} -1, & x < -2 \\ [-4; -1], & x = -2 \\ 2x, & |x| < 2 \\ [1; 4], & x = 2 \\ 1, & x > 2 \end{cases}.$$

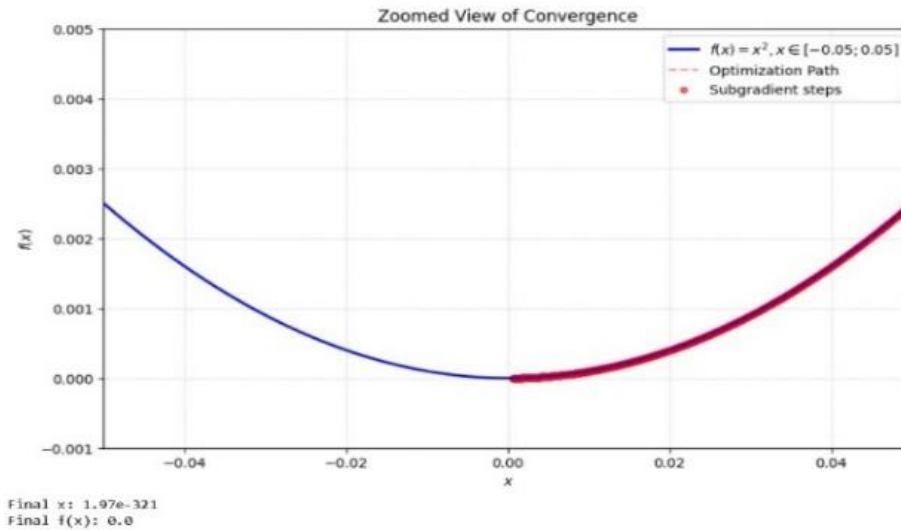
Ta chọn $x^{(1)} = 3$, độ chính xác $\eta = 0,015$ và $\varepsilon = 0,005$. Thông thường ta chỉ có thể ước lượng một chặn đủ lớn của $\|x^{(1)} - x^*\|$, chẳng hạn ta chọn $R = 5$. Ngoài ra, ta có thể chọn $G = \max_{1 \leq i \leq k} \|u^{(i)}\| = 4$. Khi đó theo Mệnh đề 3.1, ta có thể chọn bước nhảy và số bước nhảy cho thuật toán này lần lượt là

$$\alpha_k = \frac{\eta - \varepsilon}{G^2} = 6,25 \cdot 10^{-4} \quad \text{và} \quad N = \frac{R^2 G^2}{(\eta - \varepsilon)^2} = 4 \cdot 10^6.$$

Kết quả thu được cho ta thấy nghiệm tối ưu của bài toán là $x^* = 0$ với giá trị tối ưu $f^* = 0$. Ở đây số bước nhảy là rất lớn, do đó tốc độ hội tụ của thuật toán sẽ rất chậm.



Hình 1: Minh họa cho thuật toán subgradient với bước nhảy hằng số.



Hình 2: Khảo sát sự hội tụ tại lân cận của nghiệm tối ưu

Tiếp theo, ta xét thuật toán subgradient với bước nhảy độ dài hằng số. Cụ thể tại bước nhảy thứ k , biểu thức của bước nhảy sẽ là $\alpha_k = \frac{\gamma}{\|u^{(k)}\|}$ với $\gamma > 0$. Khi đó thuật toán sẽ được tóm tắt trong Bảng 2.

Bảng 2: Thuật toán subgradient với bước nhảy độ dài hằng số

Đầu vào:	Nhập γ và số bước nhảy N .
Khởi tạo:	Chọn điểm đầu $x^{(1)} \in \mathbb{R}^n$.
Các bước chính:	Với mọi $0 < k \leq N$, ta thực hiện các bước sau:
	1. Tính $u^{(k)}$. Nếu $u^{(k)} = 0$ thì $x^* = x^{(k)}$, ta dừng thuật toán;
	2. Tìm $x^{(k+1)} = x^{(k)} - \gamma \frac{u^{(k)}}{\ u^{(k)}\ }$, đặt $k := k + 1$ rồi trở lại bước 1.

Ta lại áp dụng thuật toán subgradient với bước nhảy này để giải bài toán ở Ví dụ 3.1. Bằng cách chọn $x^{(1)} = 3, \gamma = 0,01$ và số bước nhảy $N = 400$, ta cũng kết luận được nghiệm tối ưu của bài toán là $x^* = 0$ với giá trị tối ưu $f^* = 0$.

TÀI LIỆU THAM KHẢO

- [1] S. Boyd, *Subgradient Methods*, Notes for EE364b, Stanford University, May 2014.
- [2] N. Shor, *Minimization Methods for Non-differentiable Functions*. Berlin, Germany: Springer-Verlag, 1985.

- [3] B.T. Polyak, *Introduction to Optimization*, Optimization Software, Inc., Publications Division, 1987.
- [4] R. M. Freund, *Subgradient Optimization, Generalized Programming, and Nonconvex Duality*, MIT Open Course Ware, Massachusetts Institute of Technology, 2004.
- [5] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal and C. A. Sagastizábal, *Numerical Optimization: Theoretical and Practical Aspects*, 2nd ed. Berlin, Germany: Springer-Verlag, 2006.
- [6] A. Scagliotti and P.C. Franzone, *A Subgradient Method with Constant Step-Size for ℓ_1 -Composite Optimization*, Bollettino dell'Unione Matematica Italiana, vol. 17, no. 3, 2024.
- [7] X. Li, L. Zhao, D. Zhu and A. M. So, *Revisiting Subgradient Method: Complexity and Convergence Beyond Lipschitz Continuity*, Vietnam Journal of Mathematics, vol. 52, no. 3, 2024.
- [8] F. Clarke, *Optimization and Nonsmooth Analysis*. John Wiley & Sons, 1983.
- [9] C. K. W. Yu, Y. Hu, X. Yang and S. K. Choy, *Abstract convergence theorem for quasi-convex optimization problems with applications*, Optimization, vol. 68, no. 7, 2019.
- [10] Y. Nesterov, *Lectures on Convex Optimization*, 2nd ed. Switzerland: Springer, 2018.
- [11] A. Beck, *First-Order Methods in Optimization*. Philadelphia, SIAM, 2017.
- [12] Đặng Đức Trọng, Phạm Hoàng Quân, Đặng Hoàng Tâm và Đinh Ngọc Thanh, *Giáo trình Giải tích hàm*. NXB Đại học quốc gia TP.HCM, 2011.
- [13] Tạ Quang Sơn, *Giáo trình Giải tích lồi và Tối ưu*. NXB Giáo dục Việt Nam, 2018.
- [14] B. Lafferriere, G. Lafferriere and N. M. Nam, *Introduction to Mathematical Analysis*. Portland State University Library, 2015.

Ngày nhận bài: 20/8/2025

Ngày chấp nhận đăng: 30/6/2026