TỐI ƯU KHÔNG CÓ RÀNG BUỘC

THUẬT TOÁN

Khoa công nghệ thông tin
Đại học PHENIKAA

Hà Nội - 2023

1. Phương pháp gradient

2. Phương pháp Newton

3. Phương pháp hướng gradient liên hợp

Phương pháp gradient

Giải bài toán tối ưu

Cho $f : \mathbb{R}^n \to \mathbb{R}$ là hàm số khả vi liên tục.

Xét bài toán tối ưu không có ràng buộc

$$\min\{f(x) : x \in \mathbb{R}^n\}.$$

- Giải phương trình $\nabla f(x) = 0$ tìm điểm dừng
- Nếu f là hàm lồi thì điểm dừng là nghiệm của bài toán.

Giải bài toán tối ưu

Cho $f : \mathbb{R}^n \to \mathbb{R}$ là hàm số khả vi liên tục.
Xét bài toán tối ưu không có ràng buộc

$$\min\{f(x) : x \in \mathbb{R}^n\}.$$

- Giải phương trình $\nabla f(x) = 0$ tìm điểm dừng
- Nếu f là hàm lồi thì điểm dừng là nghiệm của bài toán.

Giải bài toán tối ưu

Cho $f : \mathbb{R}^n \to \mathbb{R}$ là hàm số khả vi liên tục.
Xét bài toán tối ưu không có ràng buộc

$$\min\{f(x) : x \in \mathbb{R}^n\}.$$

- Giải phương trình $\nabla f(x) = 0$ tìm điểm dừng
- Nếu f là hàm lồi thì điểm dừng là nghiệm của bài toán.

Giải bài toán tối ưu

Cho $f : \mathbb{R}^n \to \mathbb{R}$ là hàm số khả vi liên tục.
Xét bài toán tối ưu không có ràng buộc

$$\min\{f(x) : x \in \mathbb{R}^n\}.$$

- Giải phương trình $\nabla f(x) = 0$ tìm điểm dừng
- Nếu f là hàm lồi thì điểm dừng là nghiệm của bài toán.

Thuật toán lặp đi tìm điểm dừng

$$x_{k+1} = x_k + t_k d_k$$

- d_k: hướng (direction)
- t_k: bước lặp (stepsize)

Hướng giảm I

Definition

Cho $f : \mathbb{R}^n \to \mathbb{R}$ là hàm số khả vi liên tục. Một véctơ $d \in \mathbb{R}^n$ được gọi là một hướng giảm *(descent direction)* của f tại x nếu đạo hàm theo hướng $f'(x; d) < 0$, tức là

$$f'(x; d) = \nabla f(x)^T d < 0.$$

Tính chất của hướng giảm

Lemma

Cho f là là hàm số khả vi liên tục trên $\mathbb{R}^n$ và $x \in \mathbb{R}^n$. Giả sử rằng d là một hướng giảm của f tại x. Khi đó, tồn tại $\epsilon > 0$ sao cho

$$f(x + td) < f(x)$$

với mọi $t \in (0, \epsilon]$.

A general descent directions method

Khởi tạo: Chọn $x_0 \in \mathbb{R}^n$ bất kỳ

Ở mỗi bước: $k = 0, 1, 2, \ldots$

(a) Chọn một hướng giảm (descent direction) d_k.

(b) Tìm bước lặp (stepsize) t_k sao cho

$$f(x_k + t_k d_k) < f(x_k).$$

(c) Đặt $x_{k+1} = x_k + t_k d_k$.

(d) Nếu tiêu chuẩn dừng thỏa mãn thì STOP và x_{k+1} là **output**

Phương pháp hướng giảm

A general descent directions method

Khởi tạo: Chọn $x_0 \in \mathbb{R}^n$ bất kỳ

Ở mỗi bước: $k = 0, 1, 2, \ldots$

(a) Chọn một hướng giảm (descent direction) d_k.

(b) Tìm bước lặp (stepsize) t_k sao cho

$$f(x_k + t_k d_k) < f(x_k).$$

(c) Đặt $x_{k+1} = x_k + t_k d_k$.

(d) Nếu tiêu chuẩn dừng thỏa mãn thì STOP và x_{k+1} là **output**

Phương pháp hướng giảm

A general descent directions method

Khởi tạo: Chọn $x_0 \in \mathbb{R}^n$ bất kỳ

Ở mỗi bước: $k = 0, 1, 2, \ldots$

(a) Chọn một hướng giảm (descent direction) d_k.

(b) Tìm bước lặp (stepsize) t_k sao cho

$$f(x_k + t_k d_k) < f(x_k).$$

(c) Đặt $x_{k+1} = x_k + t_k d_k$.

(d) Nếu tiêu chuẩn dừng thỏa mãn thì STOP và x_{k+1} là

output

A general descent directions method

Khởi tạo: Chọn $x_0 \in \mathbb{R}^n$ bất kỳ

Ở mỗi bước: $k = 0, 1, 2, \ldots$

(a) Chọn một hướng giảm (descent direction) d_k.

(b) Tìm bước lặp (stepsize) t_k sao cho

$$f(x_k + t_k d_k) < f(x_k).$$

(c) Đặt $x_{k+1} = x_k + t_k d_k$.

(d) Nếu tiêu chuẩn dừng thỏa mãn thì STOP và x_{k+1} là **output**

Phương pháp hướng giảm

A general descent directions method

Khởi tạo: Chọn $x_0 \in \mathbb{R}^n$ bất kỳ

Ở mỗi bước: $k = 0, 1, 2, \ldots$

(a) Chọn một hướng giảm (descent direction) d_k.

(b) Tìm bước lặp (stepsize) t_k sao cho

$$f(x_k + t_k d_k) < f(x_k).$$

(c) Đặt $x_{k+1} = x_k + t_k d_k$.

(d) Nếu tiêu chuẩn dừng thỏa mãn thì STOP và x_{k+1} là **output**

Phương pháp hướng giảm

A general descent directions method

Khởi tạo: Chọn $x_0 \in \mathbb{R}^n$ bất kỳ

Ở mỗi bước: $k = 0, 1, 2, \ldots$

(a) Chọn một hướng giảm (descent direction) d_k.

(b) Tìm bước lặp (stepsize) t_k sao cho

$$f(x_k + t_k d_k) < f(x_k).$$

(c) Đặt $x_{k+1} = x_k + t_k d_k$.

(d) Nếu tiêu chuẩn dừng thỏa mãn thì STOP và x_{k+1} là **output**

Phương pháp hướng giảm

A general descent directions method

Khởi tạo: Chọn $x_0 \in \mathbb{R}^n$ bất kỳ

Ở mỗi bước: $k = 0, 1, 2, \ldots$

(a) Chọn một hướng giảm (descent direction) d_k.

(b) Tìm bước lặp (stepsize) t_k sao cho

$$f(x_k + t_k d_k) < f(x_k).$$

(c) Đặt $x_{k+1} = x_k + t_k d_k$.

(d) Nếu tiêu chuẩn dừng thỏa mãn thì STOP và x_{k+1} là **output**

??????????????

- Bước lặp (stepsize) chọn như thế nào?
- Tiêu chuẩn dừng thuật toán là gì?

??????????????

- Bước lặp (stepsize) chọn như thế nào?
- Tiêu chuẩn dừng thuật toán là gì?

Chọn bước lặp (stepsize)

- Bước lặp hằng $t_k = t$ với mọi k
- Bước lặp theo *"exact line search"* t_k là cực tiểu dọc theo tia $x_k + td_k$:

$$t_k = \operatorname{argmin}_{t \geq 0} f(x_k + td_k)$$

- Backtracking

Chọn bước lặp (stepsize)

- Bước lặp hằng $t_k = t$ với mọi k
- Bước lặp theo *"exact line search"* t_k là cực tiểu dọc theo tia $x_k + td_k$:

$$t_k = \text{argmin}_{t \geqslant 0} f(x_k + td_k)$$

- Backtracking

Chọn bước lặp (stepsize)

- Bước lặp hằng $t_k = t$ với mọi k
- Bước lặp theo *"exact line search"* t_k là cực tiểu dọc theo tia $x_k + td_k$:

$$t_k = \operatorname{argmin}_{t \geqslant 0} f(x_k + td_k)$$

- Backtracking

- Tiêu chuẩn dừng thuật toán:

$$\|\nabla f(x_{k+1})\| < \epsilon$$

ϵ là một số dương đủ bé được chọn trước (input) (Thông thường, cho $\epsilon = 10^{-6}$ hoặc $\epsilon = 10^{-5}$)

Nhắc lại: Chuẩn của một véctơ $x = (a_1, a_2, \ldots, a_n)$ trong $\mathbb{R}^n$ được định nghĩa là

$$\|x\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}.$$

- Tiêu chuẩn dừng thuật toán:

$$\|\nabla f(x_{k+1})\| < \epsilon$$

ϵ là một số dương đủ bé được chọn trước (input) (Thông thường, cho $\epsilon = 10^{-6}$ hoặc $\epsilon = 10^{-5}$)

Nhắc lại: Chuẩn của một véctơ $x = (a_1, a_2, \ldots, a_n)$ trong $\mathbb{R}^n$ được định nghĩa là

$$\|x\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}.$$

- Tiêu chuẩn dừng thuật toán:

$$\|\nabla f(x_{k+1})\| < \epsilon$$

ϵ là một số dương đủ bé được chọn trước (input) (Thông thường, cho $\epsilon = 10^{-6}$ hoặc $\epsilon = 10^{-5}$)

Nhắc lại: Chuẩn của một véctơ $x = (a_1, a_2, \ldots, a_n)$ trong $\mathbb{R}^n$ được định nghĩa là

$$\|x\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}.$$

Hướng giảm: ngược hướng gradient

$d_k = -\nabla f(x_k)$ là một hướng giảm tại x_k nếu $\nabla f(x_k) \neq 0$.

Thật vậy

$$f'(x_k; -\nabla f(x_k)) = -\nabla f(x_k)^T \nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0.$$

Hướng giảm: ngược hướng gradient

$d_k = -\nabla f(x_k)$ là một hướng giảm tại x_k nếu $\nabla f(x_k) \neq 0$.
Thật vậy

$$f'(x_k; -\nabla f(x_k)) = -\nabla f(x_k)^T \nabla f(x_k) = -\|\nabla f(x_k)\|^2 < 0.$$

Thuật toán gradient (gradient method) I

Input: $\epsilon > 0$ - tham số dung sai (*tolerance parameter*) .

Khởi tạo: $x_0 \in \mathbb{R}^n$ bất kỳ.

Với $k = 0, 1, 2, \ldots,$ thực hiện các bước sau:

- Chọn bước lặp (*stepsize*) t_k là hằng số, backtracking, hay line search bằng một chương trình trên hàm số

$$g(t) = f(x_k - t_k \nabla f(x_k)).$$

- Đặt $x_{k+l} = x_k - t_k \nabla f(x_k)$.

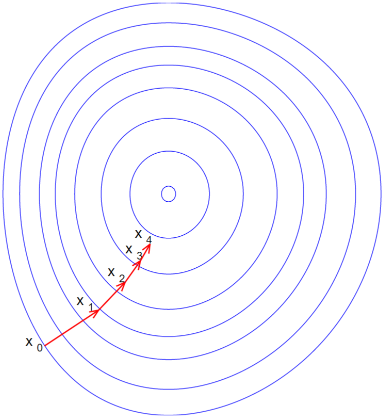
Thuật toán gradient (gradient method) II

- Nếu $\|\nabla f(x_{k+1})\| < \epsilon$ thì **STOP**, và x_{k+1} là output (nghiệm xấp xỉ)

Thuật toán gradient (gradient method) III

Bước lặp	x_k	$\|\nabla f(x_k)\|$
k=0	$\ldots$	$\ldots$
k=1	$\ldots$	$\ldots$
k=2	$\ldots$	$\ldots$
k=3	$\ldots$	$\ldots$
$\vdots$	$\vdots$	$\vdots$
$\ldots$	$\ldots$	$\ldots$

Thuật toán gradient (gradient method) IV



Hình: Gradient Method

Thuật toán gradient (gradient method) V

Ví dụ: Xây dựng dãy lặp theo phương pháp gradient cho bài toán

$$\{\min f(x, y) = x^2 + 2y^2 : (x, y) \in \mathbb{R}^2\}$$

Điểm bắt đầu: $(x_0, y_0) = (2, 1)$
Stepsize: $t = 0.1$

Phương pháp Newton

Phương pháp Newton I

Phương pháp Newton cũng là một phương pháp hướng giảm, trong đó stepsize

$$t_k = \frac{1}{\|\nabla^2 f(x_k)\|}$$

- Điểm bắt đầu: $x_0 \in \mathbb{R}^n$
- $x_{k+1} = x_k - (\|\nabla^2 f(x_k)\|)^{-1} \nabla f(x_k)$

Phương pháp Newton II

Ví dụ: Xây dựng dãy lặp theo phương pháp Newton cho bài toán

$$\{\min f(x,y) = x^2 + 2y^2 : (x,y) \in \mathbb{R}^2\}$$

Điểm bắt đầu: $(x_0, y_0) = (2, 1)$

Phương pháp hướng gradient liên hợp