ĐẠI HỌC HUẾ
TRƯỜNG ĐẠI HỌC KHOA HỌC

HUỲNH THẾ PHÙNG

Bài giảng

CÁC PHƯƠNG PHÁP TỐI ƯU

(Dành cho Cao học)

Huế - 2021

Nội dung học phần: Gồm ba chương

Chương 1. Cơ sở toán học

1.1. Khai triển Taylor

1.2. Đường mức, mặt mức

1.3. Bài toán tìm cực trị

Chương 2. Tìm kiếm theo tia

2.1. Phương pháp Nhát cắt vàng, Phương pháp Fibonacci

2.2. Các phương pháp nội suy

Chương 3. Các phương pháp sử dụng Gradient

3.1. Phương pháp giảm nhanh nhất

3.2. Các phương pháp Gradient cải tiến

3.3. Phương pháp Newton

Hàm nhiều biến

Trong giáo trình này ta luôn kí hiệu f là hàm n biến, tức là

$$f : \mathbb{R}^n \to \mathbb{R},$$

$$x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mapsto f(x) = f(x_1, x_2, \ldots, x_n) \in \mathbb{R}.$$

Khi f là hàm 2 hoặc 3 biến ta viết đơn giản là $f(x, y)$ hay $f(x, y, z)$. Chẳng hạn

$$f(x, y) = 2x^3y^2 - 7xe^y; \quad f(x, y, z) = x^2(y + z) + 5y^3xz.$$

Đạo hàm riêng - Gradient

Hàm n biến $f(x) = f(x_1, x_2, \dots, x_n)$ có n đạo hàm riêng theo mỗi biến. Để tính đạo hàm riêng của f theo một biến nào đó ta chỉ cần xem các biến còn lại là hằng số. Ví dụ, hàm hai biến $f(x, y) = 2x^3y^2 - 7xe^y$ có hai đạo hàm riêng theo x và theo y:

$$\frac{\partial f}{\partial x}(x, y) = 6x^2y^2 - 7e^y; \quad \frac{\partial f}{\partial y}(x, y) = 4x^3y - 7xe^y.$$

Gradient của f tại một điểm chính là vec-tơ gồm các đạo hàm riêng:

$$\nabla f(x, y) = (6x^2y^2 - 7e^y, 4x^3y - 7xe^y). \tag{1.1}$$

Tương tự, với hàm ba biến $g(x, y, z) = x^2(y + z) + 5y^3xz$, ta có

$$\nabla g(x, y, z) = (2x(y + z) + 5y^3z, x^2 + 15y^2xz, x^2 + 5y^3x). \tag{1.2}$$

Tổng quát, $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \dots, \frac{\partial f}{\partial x_n}(x)\right)$.

Đạo hàm riêng cấp 2 - Hessian

Hàm n biến $f(x) = f(x_1, x_2, \ldots, x_n)$ lại có n^2 đạo hàm riêng cấp 2, lập thành một ma trận vuông cấp n, gọi là Hessian của f. Chẳng hạn, hàm 2 biến $f(x, y)$ có Hessian là ma trận vuông cấp 2

$$\begin{pmatrix} \frac{\partial^2 f}{\partial x^2}(x, y) & \frac{\partial^2 f}{\partial y \partial x}(x, y) \\ \frac{\partial^2 f}{\partial x \partial y}(x, y) & \frac{\partial^2 f}{\partial y^2}(x, y) \end{pmatrix}$$

Ví dụ 1.1 ($f(x, y) = 2x^3y^2 - 7xe^y$; $g(x, y, z) = x^2(y + z) + 5y^3xz$)

$$\nabla^2 f(x, y) = \begin{pmatrix} 12xy^2 & 12x^2y - 7e^y \\ 12x^2y - 7e^y & 4x^3 - 7xe^y \end{pmatrix}, \tag{1.3}$$

$$\nabla^2 g(x, y, z) = \begin{pmatrix} 2(y + z) & 2x + 15y^2z & 2x + 5y^3 \\ 2x + 15y^2z & 30yxz & 15y^2x \\ 2x + 5y^3 & 15y^2x & 0 \end{pmatrix} \tag{1.4}$$

Khai triển Taylor cấp 1

$$f(x) = f(x^0) + \langle \nabla f(x^0), x - x^0 \rangle + o(\|x - x_0\|), \; x \in \mathbb{R}^n,$$

$$f(x) \simeq f(x^0) + \langle \nabla f(x^0), x - x^0 \rangle.$$

Khai triển Taylor cấp 2

$$f(x) = f(x^0) + \langle \nabla f(x^0), x - x^0 \rangle + \frac{1}{2}(x - x^0)^T \nabla^2 f(x^0)(x - x^0)$$

$$+ o(\|x - x^0\|^2), \; x \in \mathbb{R}^n.$$

$$f(x) \simeq f(x^0) + \langle \nabla f(x^0), x - x^0 \rangle + \frac{1}{2}(x - x^0)^T \nabla^2 f(x^0)(x - x^0).$$

Khai triển Taylor cấp 1

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + o(|x - x_0|), \quad x \in \mathbb{R}.$$

Dạng xấp xỉ: $f(x) \simeq f(x_0) + f'(x_0)(x - x_0)$.

$$\begin{aligned} f(x, y) &\simeq f(x_0, y_0) + \langle \nabla f(x_0, y_0), (x, y) - (x_0, y_0) \rangle \\ &\simeq f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0). \end{aligned}$$

$$\begin{aligned} f(x, y, z) &\simeq f(x_0, y_0, z_0) + \langle \nabla f(x_0, y_0, z_0), (x, y, z) - (x_0, y_0, z_0) \rangle \\ &\simeq f(x_0, y_0, z_0) + \frac{\partial f}{\partial x}(x_0, y_0, z_0)(x - x_0) \\ &+ \frac{\partial f}{\partial y}(x_0, y_0, z_0)(y - y_0) + \frac{\partial f}{\partial z}(x_0, y_0, z_0)(z - z_0). \end{aligned}$$

$$f(x) \simeq f(x^0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^0)(x_i - x_i^0); \; x \in \mathbb{R}^n.$$

Khai triển Taylor cấp 2

$$f(x) \simeq f(x_0) + f'(x_0)(x - x_0) + \tfrac{1}{2} f''(x_0)(x - x_0)^2; \quad x \in \mathbb{R}.$$

$$\begin{aligned}
f(x, y) &\simeq f(x_0, y_0) + \langle \nabla f(x_0, y_0), (x, y) - (x_0, y_0) \rangle \\
&\quad + \frac{1}{2}(x - x_0, y - y_0) \nabla^2 f(x_0, y_0) \begin{pmatrix} x - x_0 \\ y - y_0 \end{pmatrix} \\
&\simeq f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial f}{\partial y}(x_0, y_0)(y - y_0) \\
&\quad + \frac{1}{2}\Big[\frac{\partial^2 f}{\partial x^2}(x_0, y_0)(x - x_0)^2 + \frac{\partial^2 f}{\partial y^2}(x_0, y_0)(y - y_0)^2 \\
&\quad + 2\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)(x - x_0)(y - y_0)\Big].
\end{aligned}$$

Khai triển Taylor cấp 2

$$
\begin{aligned}
f(x,y,z) &\simeq f(x_0,y_0,z_0) + \langle \nabla f(x_0,y_0,z_0), (x,y,z)-(x_0,y_0,z_0)\rangle \\
&\qquad + \frac{1}{2}(x-x_0, y-y_0, z-z_0)\nabla^2 f(x_0,y_0,z_0)\begin{pmatrix} x-x_0 \\ y-y_0 \\ z-z_0 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&\simeq f(x_0,y_0,z_0) + \frac{\partial f}{\partial x}(x_0,y_0,z_0)(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0,z_0)(y-y_0) \\
&+ \frac{\partial f}{\partial z}(x_0,y_0,z_0)(z-z_0) + \frac{1}{2}\Big[\frac{\partial^2 f}{\partial x^2}(x_0,y_0,z_0)(x-x_0)^2 \\
&+ \frac{\partial^2 f}{\partial y^2}(x_0,y_0,z_0)(y-y_0)^2 + 2\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0,z_0)(x-x_0)(y-y_0) \\
&+ \frac{\partial^2 f}{\partial z^2}(x_0,y_0,z_0)(z-z_0)^2 + 2\frac{\partial^2 f}{\partial y \partial z}(x_0,y_0,z_0)(y-y_0)(z-z_0) \\
&+ 2\frac{\partial^2 f}{\partial z \partial x}(x_0,y_0,z_0)(z-z_0)(x-x_0)\Big].
\end{aligned}
$$

Ví dụ 1.2 ($f(x,y) = 2x^3y^2 - 7xe^y$, tại $(1,0)$)

$f(1,0) = -7$, $\nabla f(1,0) = (-7,-7)$, $\nabla^2 f(1,0) = \begin{pmatrix} 0 & -7 \\ -7 & -3 \end{pmatrix}$. Do đó khai triển Taylor cấp 1 và cấp 2 của f tại $(1,0)$ lần lượt là:

$$\begin{aligned} f(x,y) &\simeq f(1,0) + \langle \nabla f(1,0), (x,y) - (1,0) \rangle \\ &\simeq -7 + \langle (-7,-7), (x-1,y) \rangle = -7x - 7y. \\ f(x,y) &\simeq -7x - 7y + \frac{1}{2}(x-1,y) \begin{pmatrix} 0 & -7 \\ -7 & -3 \end{pmatrix} \begin{pmatrix} x-1 \\ y \end{pmatrix} \\ &\simeq -7x - 7y + \frac{1}{2}\Big[-3y^2 - 14(x-1)y \Big]. \end{aligned}$$

Ví dụ 1.3 ($g(x,y,z) = x^2(y+z) + 5y^3xz$, tại $(1,1,0)$)

$$g(1,1,0) = 1, \nabla g(1,1,0) = (2,1,6), \nabla^2 g(1,1,0) = \begin{pmatrix} 2 & 2 & 7 \\ 2 & 0 & 15 \\ 7 & 15 & 0 \end{pmatrix}$$

Khai triển Taylor cấp 1 và 2 của g tại $(1,1,0)$ là

$$g(x,y,z) \simeq 1 + \langle (2,1,6), (x-1, y-1, z) \rangle = 2x + y + 6z - 2$$

$$g(x,y,z) \simeq 2x+y+6z-2 + \frac{1}{2}(x-1, y-1, z) \begin{pmatrix} 2 & 2 & 7 \\ 2 & 0 & 15 \\ 7 & 15 & 0 \end{pmatrix} \begin{pmatrix} x-1 \\ y-1 \\ z \end{pmatrix}$$

$$\simeq 2x + y + 6z - 2 + \frac{1}{2}\Big[2(x-1)^2 + 4(x-1)(y-1) + 14(x-1)z + 30(y-1)z\Big].$$

Mặt mức

- Cho hàm $f : \mathbb{R}^n \to \mathbb{R}$. Với mỗi giá trị $\alpha \in \mathbb{R}$ ta gọi tập hợp

$$C(f;\alpha) := \{x \in \mathbb{R}^n \mid f(x) = \alpha\}$$

là Mặt mức của hàm f ứng với mức α.
- Nếu f là hàm hai biến thì $C(f;\alpha)$ là một đường cong trong mặt phẳng nên ta cũng gọi nó là Đường mức.

Các nửa không gian

Mặt mức (hay Đường mức) $C(f;\alpha)$ chia không gian ra làm hai nửa, gọi là các nửa không gian:

$$H^+(f;\alpha) = \{x \in \mathbb{R}^n \mid f(x) \geq \alpha\},$$

$$H^-(f;\alpha) = \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}.$$

Ví dụ 1.4 (Mặt mức - Đường mức - Nửa không gian)

- Hàm 2 biến $f(x, y) = x^2 + y^2$ có đường mức

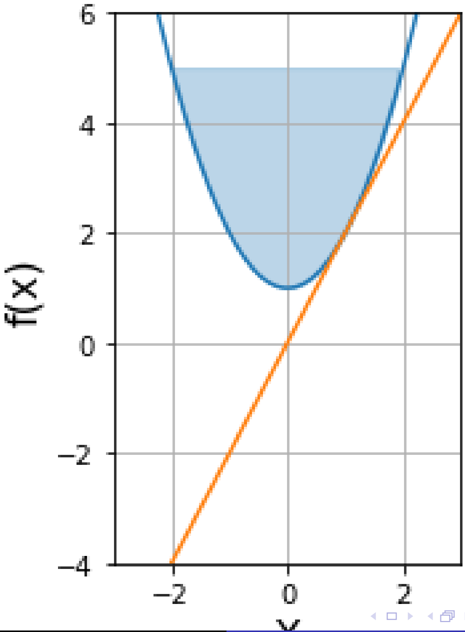
$$C(f; 4) = \{(x, y) \mid x^2 + y^2 = 4\}$$

là đường tròn tâm $O(0, 0)$ bán kính 2, chia mặt phẳng làm 2 nửa không gian, với $H^-(f; 4)$ là hình tròn và $H^+(f; 4)$ là phần bên ngoài hình tròn.

- Hàm $f(x, y) = y - x^2$ có đường mức

$$C(f; 1) = \{(x, y) \mid y - x^2 = 1\}$$

là đường parabol $y = x^2 + 1$. Đường này chia mặt phẳng ra làm hai nửa không gian, với $H^-(f; 1)$ là phần mặt phẳng nằm phía dưới parabol, $H^+(f; 1)$ là phần mặt phẳng nằm phía trên parabol.


6
4
2
0
−2
−4
f(x)
−2
0
2
x

Siêu phẳng

Cho vec-tơ $a \neq 0$. Các siêu phẳng nhận a làm vec-tơ pháp có dạng

$$N = \{x \in \mathbb{R}^n \mid \langle a, x \rangle = b\}.$$

Siêu phẳng nhận a làm vec-tơ pháp và đi qua điểm x^0 là

$$N = \{x \in \mathbb{R}^n \mid \langle a, x - x^0 \rangle = 0\}.$$

Ví dụ 1.5

Trong $\mathbb{R}^2$, siêu phẳng nhận $a = (1, -2)$ làm vec-tơ pháp và đi qua $(5, 2)$ là đường thẳng $\langle (1, -2), (x - 5, y - 2) \rangle = 0 \Leftrightarrow x - 2y - 1 = 0$.

Ví dụ 1.6

Trong $\mathbb{R}^3$, siêu phẳng nhận $a = (2, -1, 3)$ làm vec-tơ pháp và đi qua điểm $(0, 5, -4)$ là mặt phẳng

$$\langle (2, -1, 3)(x, y - 5, z + 4) \rangle = 0 \Leftrightarrow 2x - y + 3z + 17 = 0.$$

Siêu phẳng tiếp xúc với mặt mức

Cho hàm n biến f, $x^0 \in \mathbb{R}^n$ và $f(x^0) = \alpha$. Lúc đó, $x^0 \in C(f;\alpha)$, và siêu phẳng tiếp xúc với $C(f;\alpha)$ tại x^0 nhận $\nabla f(x^0)$ làm vec-tơ pháp. Cụ thể, siêu phẳng tiếp xúc với $C(f;\alpha)$ tại x^0 có phương trình

$$\langle \nabla f(x^0), x - x^0 \rangle = 0.$$

Vec-tơ $\nabla f(x^0)$ hướng vào miền $H^+(f;\alpha)$

Ngoài ra, vec-tơ $\nabla f(x^0)$ hướng vào miền $H^+(f;\alpha)$. Cụ thể,

(a) Nếu $v \in \mathbb{R}^n$ sao cho $\langle \nabla f(x^0), v \rangle > 0$ thì với $t > 0$ đủ bé, ta có

$$f(x^0 + tv) > f(x^0). \ (*)$$

(b) Nếu $v \in \mathbb{R}^n$ sao cho $\langle \nabla f(x^0), v \rangle < 0$ thì với $t > 0$ đủ bé, ta có

$$f(x^0 + tv) < f(x^0). \ (**)$$

Ví dụ 1.7 ($f(x, y) = x^2 + y^2$)

Vì $f(\sqrt{3}, 1) = 4$ nên $(\sqrt{3}, 1) \in C(f; 4)$. Vec-tơ $\nabla f(\sqrt{3}, 1) = (2\sqrt{3}, 2)$ vuông góc với $C(f; 4)$ và hướng ra ngoài (tức là hướng vào miền $H^+(f; 4)$). Đường thẳng tiếp xúc với đường tròn $C(f; 4)$ tại $(\sqrt{3}, 1)$ có phương trình:

$$\langle (2\sqrt{3}, 2), (x - \sqrt{3}, y - 1) \rangle = 0 \Leftrightarrow \sqrt{3}x + y - 4 = 0.$$

Ví dụ 1.8 ($f(x, y) = y - x^2$)

Vì $f(1, 2) = 1$ nên $(1, 2) \in C(f; 1)$. Vec-tơ $\nabla f(1, 2) = (-2, 1)$ vuông góc với $C(f; 1)$ và hướng lên trên (tức là hướng vào miền $H^+(f; 1)$). Đường thẳng tiếp xúc với $C(f; 1)$ tại $(1, 2)$ có phương trình:

$$\langle (-2, 1), (x - 1, y - 2) \rangle = 0 \Leftrightarrow -2x + y = 0.$$

Hướng tăng, Hướng giảm

• Một vec-tơ $v \in \mathbb{R}^n$ thoả mãn $(*)$ được gọi là một hướng tăng, và thoả mãn $(**)$ được gọi là một hướng giảm của hàm f tại x^0.
• Như vậy, nếu vec-tơ v lập với $\nabla f(x^0)$ một góc nhọn thì nó là hướng tăng, còn nếu lập với $\nabla f(x^0)$ một góc tù thì là hướng giảm của f tại x^0.

Ý nghĩa hình học

Cũng từ các công thức $(*)$ và $(**)$ ta thấy:
• Nếu v là một hướng tăng của f tại x^0 thì, với $t > 0$ đủ bé ta có

$$x^0 + tv \in H^+(f;\alpha).$$

• Nếu v là một hướng giảm của f tại x^0 thì, với $t > 0$ đủ bé ta có

$$x^0 + tv \in H^-(f;\alpha).$$

Cực tiểu địa phương

Một điểm X^* được gọi là cực tiểu địa phương của f, nếu tồn tại số dương ϵ sao cho

$$f(x^*) \leq f(x); \ \forall x \in B(x^*; \epsilon).$$

Điều kiện cần

Nếu x^* là một điểm cực tiểu địa phương của f thì

$$\nabla f(x^*) = 0.$$

Điều kiện đủ

Nếu $\nabla f(x^*) = 0$ và f là hàm lồi (hoặc $\nabla^2 f(x)$ nửa xác định dương trong một lân cận của x^*, hoặc $\nabla^2 f(x^*)$ xác định dương), thì x^* là một điểm cực tiểu địa phương của f.

Tư tưởng thuật toán

Xuất phát từ một điểm x_0 ta sẽ sinh ra các điểm x_1, x_2, ... sao cho dãy điểm (x_k) sinh ra hội tụ về nghiệm. Tại mỗi bước thứ k trước tiên ta cần xác định một hướng giảm tại đó, tức là tìm vec-tơ d_k sao cho $\langle \nabla f(x_k), d_k \rangle < 0$. Sau đó ta tìm một bước giảm thích hợp $\alpha_k > 0$ sao cho $f(x_k + \alpha_k d_k) < f(x_k)$. Lúc đó ta đặt $x_{k+1} := x_k + \alpha_k d_k$.

Thuật toán tổng quát tìm điểm cực tiểu

Bước 0. Lấy $x_0 \in \mathbb{R}^n$, chọn sai số $\epsilon > 0$, đặt $k := 0$.
Bước 1. Nếu $\|\nabla f(x_k)\| < \epsilon$, thì dừng. Ngược lại sang Bước 2.
Bước 2. Tìm hướng giảm d_k: $\langle \nabla f(x_k), d_k \rangle < 0$.
Bước 3. Tìm bước giảm $\alpha_k > 0$ sao cho $f(x_k + \alpha_k d_k) < f(x_k)$.
Bước 4. Đặt $x_{k+1} := x_k + \alpha_k d_k$, $k := k + 1$.
Bước 5. Nếu $\|x_{k+1} - x_k\| < \epsilon$ hoặc $|f(x_{k+1}) - f(x_k)| < \epsilon$, thì dừng. Ngược lại sang Bước 1.

Nội dung

Trong thuật toán tổng quát, tại bước lặp thứ k, khi đã có x_k và hướng giảm d_k, ta cần tìm bước giảm $\alpha_k > 0$ sao cho:

$$f(x_k + \alpha_k d_k) < f(x_k).$$

Lí tưởng nhất là ta tìm được $\alpha_k > 0$ sao cho:

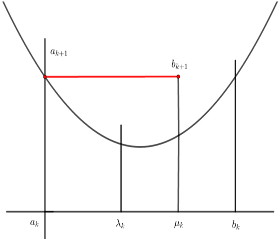
$$f(x_k + \alpha_k d_k) = \min_{\alpha > 0} f(x_k + \alpha d_k).$$

Tức là ta giải bài toán cực tiểu hàm một biến $\phi(\alpha) := f(x_k + \alpha d_k)$ trên nửa đường thẳng $(0, +\infty)$.
Đó gọi là bài toán tìm kiếm theo tia.

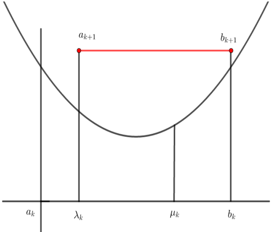
Khoảng đơn cách

Khoảng $[a, b]$ được gọi là một khoảng đơn cách của hàm ϕ nếu tồn tại $x^* \in (a, b)$ sao cho ϕ giảm trên $[a, x^*]$ và tăng trên $[x^*, b]$ Bây giờ giả sử $< \lambda < \mu < b$, ta có 2 khả năng:

- Nếu $\phi(\lambda) \leq \phi(\mu)$ thì $[a, \mu]$ là một khoảng đơn cách mới của ϕ.
- Nếu $\phi(\lambda) > \phi(\mu)$ thì $[\lambda, b]$ là một khoảng đơn cách mới của ϕ.



Trường hợp 1



Trường hợp 2

Ý tưởng các phương pháp

Xuất phát từ một khoảng đơn cách $[a_0, b_0]$ cho trước, ta chọn λ_0, μ_0 sao cho $a_0 < \lambda_0 < \mu_0 < b_0$. Lúc đó một trong hai khoảng $[a_0, \mu_0]$ và $[\lambda_0, b_0]$ sẽ là khoảng đơn cách. Ta đặt tên khoảng đó là $[a_1, b_1]$. Lại tiếp tục chọn λ_1, μ_1 sao cho $a_1 < \lambda_1 < \mu_1 < b_1$ và lại chọn ra khoảng đơn cách $[a_2, b_2]$. Cứ tiếp tục như thế, đến một bước nào đó, ta nhận được khoảng đơn cách $[a_k, b_k]$ có độ dài bé hơn một ngưỡng ϵ cho trước. Lúc này mọi điểm trong khoảng $[a_k, b_k]$ đều có thể xem là điểm cực tiểu gần đúng của hàm số, với sai số ϵ.

Cách chọn λ_k, μ_k

Trước tiên ta chọn $\tau_k \in (0.5; 1)$. Sau đó đặt

$$\lambda_k = a_k + (1 - \tau_k)(b_k - a_k); \quad \mu_k = a_k + \tau_k(b_k - a_k).$$

Lúc đó $a_k < \lambda_k < \mu_k < b_k$, và $b_k - \lambda_k = \mu_k - a_k = \tau_k(b_k - a_k)$.
Vì vậy, τ_k được gọi là độ co rút tại bước k.

Phương pháp Nhát cắt vàng

Phương pháp Nhát cắt vàng chọn $\tau_k = 0.618$ cho mọi k. Lúc đó

$$\lambda_k := a_k + 0.382(b_k - a_k),\ \mu_k := a_k + 0.618(b_k - a_k).$$

Phương pháp Fibonacci

Phương pháp Fibonacci chọn τ_k khác nhau cho các bước khác nhau, dựa vào dãy số Fibonacci (F_k) : $1; 1; 2; 3; 5; 8; 13; 21; \ldots$. Cụ thể, đầu tiên chọn n là số đủ lớn sao cho

$$\frac{b_0 - a_0}{F_{n+1}} < \epsilon.$$

Tại mỗi $k < n - 1$, chọn $\tau_k = \dfrac{F_{n-k}}{F_{n-k+1}} \Rightarrow 1 - \tau_k = \dfrac{F_{n-k-1}}{F_{n-k+1}}$. Do đó

$$\lambda_k := a_k + \frac{F_{n-k-1}}{F_{n-k+1}}(b_k - a_k),\ \mu_k := a_k + \frac{F_{n-k}}{F_{n-k+1}}(b_k - a_k).$$

Sự tiện lợi của 2 phương pháp

Bằng cách chọn đặc biệt, cả hai phương pháp đều có điểm chung là tiết kiệm trong tính toán. Cụ thể là ngoại trừ bước đầu tiên ta cần tính cả λ_0 và μ_0. Ở các bước sau ta chỉ cần tính một trong hai giá trị λ_{k+1} hoặc μ_{k+1} và tận dụng một trong hai giá trị λ_k hoặc μ_k của bước trước. Tại mỗi bước ta luôn đặt $\phi_1 = \phi(\lambda_k)$, $\phi_2 = \phi(\mu_k)$.

Trường hợp 1: Nếu $\phi_1 > \phi_2$

Ta sẽ đặt $a_{k+1} := \lambda_k$, $b_{k+1} := b_k$, $\lambda_{k+1} := \mu_k$, $\phi_1 := \phi_2$.
Tính $\mu_{k+1} := a_{k+1} + \tau_{k+1}(b_{k+1} - a_{k+1})$; $\phi_2 := \phi(\mu_{k+1})$.

Trường hợp 2: Nếu $\phi_1 \le \phi_2$

Ta sẽ đặt $a_{k+1} := a_k$, $b_{k+1} := \mu_k$, $\mu_{k+1} := \lambda_k$, $\phi_2 := \phi_1$.
Tính $\lambda_{k+1} := a_{k+1} + (1 - \tau_{k+1})(b_{k+1} - a_{k+1})$; $\phi_1 := \phi(\lambda_{k+1})$.

Bước 0. Xác định $[a_0, b_0]$, chọn $\epsilon > 0$, $k := 0$. Tính

$$\lambda_0 = a_0 + 0,382(b_0 - a_0); \mu_0 = a_0 + 0,618(b_0 - a_0);$$
$$\phi_1 := \phi(\lambda_0); \phi_2 := \phi(\mu_0).$$

Bước 1. So sánh, Nếu $\phi_1 > \phi_2$: sang B2, nếu $\phi_1 \leq \phi_2$: sang B3.

Bước 2. Nếu $b_k - \lambda_k \leq \epsilon$ Dừng với μ_k là nghiệm. Ngược lại thì đặt $a_{k+1} := \lambda_k$; $b_{k+1} := b_k$; $\lambda_{k+1} := \mu_k$; $\phi_1 := \phi_2$; Tính

$$\mu_{k+1} := a_{k+1} + 0.618(b_{k+1} - a_{k+1}); \ \phi_2 := \phi(\mu_{k+1}).$$

Sang Bước 4.

Bước 3. Nếu $\mu_k - a_k \leq \epsilon$ Dừng với λ_k là nghiệm. Ngược lại thì đặt $a_{k+1} := a_k$; $b_{k+1} := \mu_k$; $\mu_{k+1} := \lambda_k$; $\phi_2 := \phi_1$; Tính

$$\lambda_{k+1} := a_{k+1} + 0.382(b_{k+1} - a_{k+1}); \ \phi_1 := \phi(\lambda_{k+1}).$$

Sang Bước 4.

Bước 4. Đặt $k := k + 1$. Sang Bước 1.

Bước 0. Cho $[a_0, b_0]$, $\epsilon > 0$, $k := 0$. Chọn n: $\frac{1}{F_{n+1}}(b_0 - a_0) < \epsilon$.
Tính $\lambda_0 = a_0 + \frac{F_{n-1}}{F_{n+1}}(b_0 - a_0)$; $\mu_0 = a_0 + \frac{F_n}{F_{n+1}}(b_0 - a_0)$;
$\phi_1 := \phi(\lambda_0)$, $\phi_2 := \phi(\mu_0)$.

Bước 1. So sánh, nếu $\phi_1 > \phi_2$ sang B2, ngược lại sang B3.

Bước 2. Nếu $k = n - 1$ thì dừng với μ_k là nghiệm. Ngược lại đặt

$$a_{k+1} := \lambda_k;\ b_{k+1} := b_k;\ \lambda_{k+1} := \mu_k;\ \phi_1 := \phi_2.$$

Tính $\mu_{k+1} := a_{k+1} + \frac{F_{n-k}}{F_{n-k+1}}(b_{k+1} - a_{k+1})$; $\phi_2 := \phi(\mu_{k+1})$.
Sang B4.

Bước 3. Nếu $k = n - 1$ thì dừng với λ_k là nghiệm. Ngược lại đặt

$$a_{k+1} := a_k;\ b_{k+1} := \mu_k;\ \mu_{k+1} := \lambda_k;\ \phi_2 := \phi_1.$$

Tính $\lambda_{k+1} := a_{k+1} + \frac{F_{n-k-1}}{F_{n-k+1}}(b_{k+1} - a_{k+1})$; $\phi_1 := \phi(\lambda_{k+1})$.
Sang Bước 4.

Bước 4. Đặt $k := k + 1$. Sang Bước 1.

Ví dụ bằng số:

Cho hàm số $\phi(t) = e^t + e^{-t}$ trên $[-1; 1]$. Hãy tìm cực tiểu hàm ϕ bằng Phương pháp nhát cắt vàng, với ngưỡng sai số $\epsilon = 0.1$.

k	a	λ	μ	b	$\phi(\lambda)$	$\phi(\mu)$	
0	-1	-0,236	0,236	1	2,056	2, 056	$\rightarrow$ B3
1	-1	-0,528	-0,236	0,236	2,285	2,056	$\rightarrow$ B2
2	-0,528	-0,236	-0,056	0,236	2,056	2,003	$\rightarrow$ B2
3	-0,236	-0,056	0,056	0,236	2,015	2,003	$\rightarrow$ B3
4	-0,236	-0,124	-0,056	0,056	2,015	2,003	$\rightarrow$ B2
5	-0,124	-0,056	-0,013	0,056	2,003	2,000	$\rightarrow$ B2
6	-0,056	-0,013	0,013	0,056	2,000	2,000	$\rightarrow$ B3
	Vì $\mu_6 - a_6 < \epsilon$ nên dừng và $\lambda_6 = -0.013$ là nghiệm.						

Table: Tóm tắt các bước lặp của phương pháp nhát cắt vàng

Ta cũng có thể giải bài toán trên bằng ngôn ngữ lập trình Python

Ví dụ bằng số:

Cho hàm số $\phi(t) = e^t + e^{-t}$ trên $[-1; 1]$. Hãy tìm cực tiểu hàm ϕ bằng Phương pháp Fibonacci, với ngưỡng sai số $\epsilon = 0.1$.

Do độ dài khoảng đơn cách ban đầu là $b_0 - a_0 = 1 - (-1) = 2$ và $\epsilon = 0.1$ nên ta chọn $n = 6$ là vừa đủ để $\frac{b_0 - a_0}{F_{n+1}} = \frac{2}{F_7} = \frac{2}{21} < \epsilon$.

k	a	λ	μ	b	$\phi(\lambda)$	$\phi(\mu)$	
0	-1	-0,238	0,238	1	2,057	2, 057	$\rightarrow$ B3
1	-1	-0,524	-0,238	0,238	2,281	2,057	$\rightarrow$ B2
2	-0,524	-0,238	-0,048	0,238	2,057	2,002	$\rightarrow$ B2
3	-0,238	-0,048	0,048	0,238	2,002	2,002	$\rightarrow$ B3
4	-0,238	-0,143	-0,048	0,048	2,020	2,002	$\rightarrow$ B2
5	-0,143	-0,048	-0,048	0,048	2,002	2,002	$\rightarrow$ B2
Vì $k = n - 1 = 5$ nên dừng và $\mu_5 = -0,048$ là nghiệm gần đúng.							

Table: Tóm tắt các bước lặp của phương pháp Fibonacci

Ý tưởng

Phương pháp nội suy là một cách tiếp cận khác của tìm kiếm trên đường thẳng, cho ta ước tính giá trị của một hàm dựa vào một số giá trị đã biết. Trong phương pháp này, ta xấp xỉ hàm $\phi(\alpha) = f(x_k + \alpha d_k)$ bằng một đa thức bậc 2 $q(\alpha) = a\alpha^2 + b\alpha + c$ (a, b, c là các hệ số cần tìm) và tìm giá trị α^* để đa thức q đạt cực tiểu. Sau đó, ta làm giảm độ dài khoảng đơn cách bằng việc so sánh giá trị α^* vừa tìm được với các điểm đã biết. Khi đã tìm được các hệ số a, b, c của đa thức nội suy ($a > 0$) thì α^* dễ dàng được xác định bởi công thức

$$\alpha^* := -\frac{b}{2a}.$$

Ta hi vọng rằng α^* cũng chứa thông tin về điểm cực tiểu của hàm ϕ.

Nội suy với hai điểm Loại I

Giả sử ta có khoảng đơn cách $[\alpha_1, \alpha_2]$ và biết các giá trị $\phi_1 := \phi(\alpha_1)$, $\phi_2 := \phi(\alpha_2)$, $\phi_2' := \phi'(\alpha_2)$. Ta tìm hàm bậc 2: $q(\alpha) = a\alpha^2 + b\alpha + c$, thoả mãn hệ:

$$\begin{cases} q(\alpha_1) = a\alpha_1^2 + b\alpha_1 + c & = \phi(\alpha_1) = \phi_1, \\ q(\alpha_2) = a\alpha_2^2 + b\alpha_2 + c & = \phi(\alpha_2) = \phi_2, \\ q'(\alpha_2) = 2a\alpha_2 + b & = \phi'(\alpha_2) = \phi_2'. \end{cases} \tag{2.1}$$

Giải (2.1) ta được
$$\begin{cases} a & = \dfrac{\phi_1 - \phi_2 - \phi_2'(\alpha_1 - \alpha_2)}{(\alpha_1 - \alpha_2)^2} \\ b & = \phi_2' - 2\dfrac{\phi_1 - \phi_2 - \phi_2'(\alpha_1 - \alpha_2)}{(\alpha_1 - \alpha_2)^2}\alpha_2 \end{cases}$$

Do đó, điểm cực tiểu của hàm $q(\alpha)$ là:

$$\alpha^* = \frac{-b}{2a} = \alpha_2 - \frac{1}{2}\frac{(\alpha_2 - \alpha_1)\phi_2'}{\phi_2' - \frac{\phi_2 - \phi_1}{\cdots}}.$$

Nội suy với hai điểm Loại II

Với khoảng đơn cách $[\alpha_1, \alpha_2]$ và các giá trị $\phi_2 := \phi(\alpha_2)$, $\phi'_1 := \phi'(\alpha_1)$, $\phi'_2 := \phi'(\alpha_2)$. Ta tìm hàm: $q(\alpha) = a\alpha^2 + b\alpha + c$, thoả mãn:

$$\begin{cases} q(\alpha_2) = a\alpha_2^2 + b\alpha_2 + c & = \phi(\alpha_2) = \phi_2, \\ q'(\alpha_2) = 2a\alpha_2 + b & = \phi'(\alpha_2) = \phi'_2, \\ q'(\alpha_1) = 2a\alpha_1 + b & = \phi'(\alpha_1) = \phi'_1. \end{cases} \tag{2.2}$$

Tương tự Phương pháp loại I, ta giải ra a, b, c, và thu được:

$$\alpha^* = \frac{-b}{2a} = \alpha_2 - \frac{\alpha_2 - \alpha_1}{\phi'_2 - \phi'_1}\phi'_2 \tag{2.3}$$

Chú ý

Thực ra Công thức (2.3) chỉ dùng 2 giá trị ϕ'_1 và ϕ'_2 để tính α^*. Vì vậy, trong thuật toán chúng ta cũng chỉ sử dụng 2 giá trị này.

Xác định khoảng đơn cách mới

Sau khi nhận được $\alpha^* \in [\alpha_1, \alpha_2]$ ta so sánh để chọn lại khoảng đơn cách mới. Để $[\alpha_1, \alpha_2]$ là khoảng đơn cách thường thì ta phải có $\phi'(\alpha_1) < 0 < \phi'(\alpha_2)$. Do đó, nếu $\phi'(\alpha^*) < 0$ ta nên chọn lại $\alpha_1 := \alpha^*$ và giữ nguyên α_2. Ngược lại nếu $\phi'(\alpha^*) > 0$ thì ta chọn lại $\alpha_2 := \alpha^*$ và giữ nguyên α_1. Thuật toán cứ tiếp tục như thế cho đến khi độ dài khoảng đơn cách $\alpha_2 - \alpha_1$ bé hơn ngưỡng ϵ, hoặc khi gặp $\phi'(\alpha^*) \approx 0$. Tóm lại, ta có các thuật toán sau:

Bước 0. Chọn ϵ, khoảng đơn cách $[\alpha_1, \alpha_2]$. Tính $\phi_1 := \phi(\alpha_1)$, $\phi_2 := \phi(\alpha_2)$, $\phi_2' := \phi'(\alpha_2)$.

Bước 1. Tìm α^* bằng công thức:

$$\alpha^* := \alpha_2 - \frac{1}{2} \frac{(\alpha_2 - \alpha_1)\phi_2'}{\phi_2' - \dfrac{\phi_2 - \phi_1}{\alpha_2 - \alpha_1}}; \tag{2.4}$$

và tính $\phi^* := \phi(\alpha^*); \ \phi'^* := \phi'(\alpha^*)$.

Bước 2. Nếu $|\phi'^*| < \epsilon$ thì dừng và cho ra nghiệm là α^*; Ngược lại thì kiểm tra, nếu $\phi'^* < 0$ sang Bước 3; nếu $\phi'^* > 0$ sang Bước 4.

Bước 3. Đặt $\alpha_1 := \alpha^*$; $\phi_1 := \phi^*$; sang Bước 5.

Bước 4. Đặt $\alpha_2 := \alpha^*$; $\phi_2 := \phi^*$; $\phi_2' := \phi'^*$; sang Bước 5.
Bước 5. Nếu $|\alpha_2 - \alpha_1| < \epsilon$ thì dừng và cho ra nghiệm là α^*, ngược lại sang Bước 1

Thuật toán Nội suy với hai điểm - Loại II

Bước 0. Chọn ϵ, khoảng đơn cách $[\alpha_1, \alpha_2]$. Tính $\phi'_1 := \phi'(\alpha_1)$, $\phi'_2 := \phi'(\alpha_2)$.

Bước 1. Tìm α^* bằng công thức:

$$\alpha^* := \alpha_2 - \frac{\alpha_2 - \alpha_1}{\phi'_2 - \phi'_1}\phi'_2,$$

và tính
$$\phi'^* := \phi'(\alpha^*).$$

Bước 2. Nếu $|\phi'^*| < \epsilon$ thì dừng và cho ra nghiệm là α^*; Ngược lại thì kiểm tra, nếu $\phi'^* < 0$ sang Bước 3; nếu $\phi'^* > 0$ sang Bước 4.

Bước 3. Đặt $\alpha_1 := \alpha^*$; $\phi'_1 := \phi'^*$; sang Bước 5.

Bước 4. Đặt $\alpha_2 := \alpha^*$; $\phi'_2 := \phi'^*$; sang Bước 5.

Bước 5. Nếu $|\alpha_2 - \alpha_1| < \epsilon$ thì dừng và cho ra nghiệm là α^*, ngược lại sang Bước 1

Ví dụ: Tìm cực tiểu hàm $\phi(t) = 1 - te^{-t^2}$ trên khoảng đơn cách $[0, 1]$ bằng phương pháp nội suy với 2 điểm Loại I, với $\epsilon = 0, 1$.

Ta có $\phi(t) = 1 - te^{-t^2}$; $\phi'(t) = 2t^2e^{-t^2} - e^{-t^2}$.
B0: Đặt $\epsilon := 0, 1$; $\alpha_1 := 0$; $\alpha_2 := 1$; $\phi_1 := \phi(0) = 1$; $\phi_2 := \phi(1) \approx 0.6321$; $\phi_2' := \phi'(1) \approx 0.3679 > 0$.

B1: Tính $\alpha^* := \alpha_2 - \frac{1}{2}\frac{(\alpha_2 - \alpha_1)\phi_2'}{\phi_2' - \frac{\phi_2 - \phi_1}{\alpha_2 - \alpha_1}} \approx 0, 7501$;

$$\phi^* := \phi(\alpha^*) \approx 0, 5727; \ \phi'^* := \phi'(\alpha^*) \approx 0, 0713.$$

B2: Vì $|\phi'^*| \geq \epsilon$ và $\phi'^* > 0$, sang Bước 4.
B4: Đặt $\alpha_2 := \alpha^* = 0, 7501$; $\phi_2 := \phi^* = 0, 5727$; $\phi_2' := \phi'^* = 0, 0713$; sang Bước 5.
B5: Do $\alpha_2 - \alpha_1 = 0, 7501 \geq \epsilon$ nên sang Bước 1.

B1: Tính $\alpha^* \approx 0, 7083$;

$$\phi^* := \phi(\alpha^*) \approx 0, 5711; \ \phi'^* := \phi'(\alpha^*) \approx 0, 0020.$$

B2: Vì $|\phi'^*| < \epsilon$ nên dừng, nghiệm gần đúng là $\alpha^* = 0, 7083$.

Ví dụ: Tìm cực tiểu hàm $\phi(t) = 1 - te^{-t^2}$ trên khoảng đơn cách $[0,1]$ bằng phương pháp nội suy 2 điểm Loại II, với $\epsilon = 0,1$.

Ta có $\phi'(t) = 2t^2e^{-t^2} - e^{-t^2}$.

B0: Đặt $\epsilon = 0,1$; $\alpha_1 := 0$; $\alpha_2 := 1$; $\phi_1' = \phi'(0) = -1 < 0$; $\phi_2' = \phi'(1) = \dfrac{1}{e} > 0$.

B1: Tính: $\alpha^* = \alpha_2 - \dfrac{\alpha_2 - \alpha_1}{\phi_2' - \phi_1'}\phi_2' \approx 0,7310$; $\phi'^* := \phi'(\alpha^*) \approx 0,0403$.

B2: Vì $|\phi'^*| \geq \epsilon$ và $\phi'^* > 0$, sang Bước 4.

B4: Đặt $\alpha_2 := \alpha^* = 0,7310$; $\phi_2' := \phi'^* = 0,0403$; sang Bước 5.

B5: Do $\alpha_2 - \alpha_1 = 0,7310 \geq \epsilon$ nên sang Bước 1.

B1: Tính $\alpha^* \approx 0,7026$; $\phi'^* := \phi'(\alpha^*) \approx -0,0076$.

B2: Vì $|\phi'^*| < \epsilon$ nên dừng, nghiệm gần đúng là $\alpha^* = 0,7026$.

• Phương pháp giảm nhanh nhất sử dụng hướng $d_k = -\nabla f(x_k)$ làm hướng giảm. Rồi tìm cực tiểu theo tia đối với hàm $\phi(\alpha) = f(x_k + \alpha d_k)$.

• Thuật toán gradient cơ bản cũng sử dụng hướng giảm $d_k = -\nabla f(x_k)$, nhưng với bước giảm cố định α: $x_{k+1} := x_k - \alpha \nabla f(x_k)$.

• Thuật toán giảm gradient có quán tính sử dụng công thức lặp: $x_{k+1} := x_k + \Delta_k$, trong đó

$$\Delta_k := \begin{cases} -\alpha \nabla f(x_0), & k = 0, \\ \gamma \Delta_{k-1} - \alpha \nabla f(x_k), & k > 0. \end{cases}$$

Phương pháp Newton

Tại mỗi bước, khai triển Taylor của hàm f tại x_k: $f(x) \approx g(x)$. Sau đó tìm x_{k+1} chính là cực tiểu của hàm toàn phương g.

Ý tưởng: Chọn véc-tơ $-\nabla f(x_k)$ làm hướng giảm tại bước lặp thứ k khi muốn tìm cực tiểu của hàm mục tiêu f.

Phương pháp giảm nhanh nhất:

Bước 0: Chọn ngưỡng $\epsilon > 0$, $x_0 \in \mathbb{R}^n$, $k := 0$. Tính $g_0 = \nabla f(x_0)$.
Bước 1: Nếu $\|g_k\| < \epsilon$ thì dừng và cho ra nghiệm x_k. Ngược lại, sang Bước 2.
Bước 2: Tìm (chính xác hoặc không chính xác) $\alpha_k > 0$ sao cho:

$$f(x_k - \alpha_k g_k) = \min_{\alpha > 0} f(x_k - \alpha g_k)$$

Bước 3: Tính $x_{k+1} = x_k - \alpha_k g_k$ và $g_{k+1} = \nabla f(x_{k+1})$
Bước 4: Đặt $k := k + 1$, quay lại bước 1.

Ví dụ: Hãy tìm cực tiểu hàm $f(x) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1$. bằng phương pháp giảm nhanh nhất, với $x_0 = (-2; 4)$ và $\epsilon = 0, 1$.

Ta có: $g_k = \nabla f(x_k) = (3x_1 - x_2 - 2; x_2 - x_1)$.
B0: Đặt $x_0 := (-2; 4)$; $\epsilon := 0, 1$; $k := 0$. Tính $g_0 := (-12, 6)$.

B1: Do $\|g_0\| = \sqrt{180} \geq \epsilon$ sang B2.
B2: Tìm cực tiểu hàm $\phi(\alpha) = f(x_0 - \alpha g_0) = 306\alpha^2 - 180\alpha + 26$ ta nhận được $\alpha_0 = \frac{5}{17}$ với $\phi(\alpha_0) = \frac{-8}{17}$.
B3: Đặt $x_1 = x_0 - \alpha_0 g_0 = \left(\frac{26}{17}; \frac{38}{17}\right)$; $g_1 := \nabla f(x_1) = \left(\frac{6}{17}; \frac{12}{17}\right)$.
B4: Đặt $k := k + 1 = 1$ sang Bước 1.

B1: Do $\|g_1\| = 0, 789 \geq \epsilon$ sang B2.
B2: Tìm cực tiểu hàm $\phi(\alpha) = f(x_1 - \alpha g_1) = \frac{54}{289}\alpha^2 - \frac{180}{289}\alpha - \frac{8}{17}$
Ta được nghiệm $\alpha_1 = \frac{5}{3}$ với $\phi(\alpha_1) = \frac{-268}{269}$.

B3: Đặt $x_2 = x_1 - \alpha_1 g_1 = \left(\frac{16}{17}; \frac{18}{17}\right)$; $g_2 := \nabla f(x_2) = \left(\frac{-4}{17}; \frac{2}{17}\right)$.
B4: Đặt $k := k+1 = 2$ sang Bước 1.

B1: Do $\|g_2\| = 0,263 \geq \epsilon$ sang B2
B2: Tìm cực tiểu hàm $\phi(\alpha) = f(x_2 - \alpha g_2) = \frac{2}{17}\alpha^2 - \frac{20}{289}\alpha - \frac{286}{289}$, ta được nghiệm $\alpha_2 = \frac{5}{17}$ với $\phi(\alpha_2) = \frac{-4912}{4913}$.
B3: Đặt $x_3 = x_2 - \alpha_2 g_2 = \left(\frac{292}{289}; \frac{296}{289}\right)$ $g_3 := \nabla f(x_3) = \left(\frac{2}{289}; \frac{4}{289}\right)$.
B4: Đặt $k := k+1 = 3$ sang Bước 1.

B1: Do $\|g_3\| = 0,015 < \epsilon$ nên thuật toán dừng với nghiệm gần đúng là

$$x_3 = \left(\frac{292}{289}; \frac{296}{289}\right) \approx (1; 1)$$

Thuật toán giảm nhanh nhất thực ra không hiệu quả như tên gọi. Bởi nó chỉ giảm nhanh ở những bước đầu nhưng thường hội tụ chậm trong nhiều bài toán thực tế. Vì vậy người ta tìm cách cải tiến thuật toán bằng nhiều cách. Đơn giản nhất người ta xét trường hợp α cố định cho mọi bước, nghĩa là, với mọi k, ta đặt

$$x_{k+1} := x_k - \alpha g_k.$$

Thuật toán này được giải thích trực quan như sau. Giả sử ta đang ở trên một ngọn núi vào ban đêm, không nhìn thấy cảnh vật và không định hướng được, làm cách nào để có thể xuống được thung lũng một cách nhanh nhất (ở đây thung lũng chính là những điểm cực tiểu trong bài toán tối ưu). Cách đơn giản là chúng ta cảm nhận độ nghiêng của mặt đất dưới chân mình hướng nào là dốc nhất (đó chính là hướng $-\nabla f(x_k)$) thì ta bước về hướng đó và bước từng bước một (nghĩa là bước α cố định), cho đến khi ta đi đến một nơi bằng phẳng (tức là $\nabla f(x_k) = 0$) thì đấy chính là thung lũng, nơi bằng phẳng và là điểm cực tiểu của đồ thị.

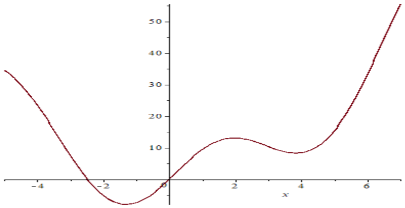
Thuật toán Gradient cơ bản

Bước 0: Chọn $\epsilon > 0$, $\alpha > 0$, $x_0 \in \mathbb{R}^n$ và đặt $k := 0$
Bước 1: Nếu $\|g_k\| < \epsilon$ dừng; ngược lại sang Bước 2.
Bước 2: Tính $x_{k+1} = x_k - \alpha g_k$
Bước 3: Đặt $k := k + 1$, sang Bước 1.

Đặc điểm

Ưu điểm là đơn giản, hội tụ khá tốt. Tuy nhiên cần lưu ý việc chọn *tốc độ học* α có ảnh hưởng rất lớn đến tốc độ hội tụ. Nếu chọn quá bé thì thuật toán hội tụ chậm, còn nếu chọn quá lớn thì có thể không hội tụ (như kiểu chúng ta xuống núi mà mỗi bước ngắn quá thì lâu đến thung lũng, còn nếu mỗi bước chúng ta nhảy xa quá thì có thể từ sườn núi này qua sườn núi kia mà không đặt chân xuống thung lũng được).

Một nhược điểm khác của thuật toán gradient cơ bản là nó thường hội tụ đến điểm cực tiểu địa phương, thay vì điểm cực tiểu toàn cục. Chúng ta có thể quan sát đồ thị hàm số $f(x) = x^2 + 10\sin(x)$



Nếu áp dụng thuật toán này với $x_0 = -4$ (tức là từ bên trái) thì sẽ hội tụ về nghiệm toàn cục, trong khi nếu xuất phát với $x_0 = 5$ (từ bên phải) thì chỉ sẽ hội tụ về nghiệm địa phương. Cụ thể, lần lượt áp dụng thuật toán với $x_0 = -4$, và $x_0 = 5$ (với $\alpha = 0.1$ và $\epsilon = 0.1$) ta được hai dãy điểm kết quả như sau:

Thuật toán gradient cơ bản

Xuất phát tại $x_0 = -4$

k	x_k	g_k
0	-4	-14.5364
1	-2.54635	-13.3728
2	-1.2090	1.1207
3	-1.3211	-0.1716
4	-1.3039	0.02863

Xuất phát tại $x_0 = 5$

k	x_k	g_k
0	5	12.8366
1	3.7163	-0.9606
2	3.8124	-0.2083
3	3.8332	-0.0354

Để khắc phục hiện tượng này người ta đề xuất Thuật toán gradient với quán tính. Để giải thích thuật ngữ này ta hình dung thuật toán tìm cực tiểu như sự di chuyển một viên bi trượt từ trên xuống dọc theo đồ thị dưới tác động của trọng lực. Thuật toán gradient cơ bản không phản ánh đúng hành động của viên bi, bởi viên bi khi di chuyển không chỉ dựa vào độ dốc tại thời điểm nó đang đứng mà còn có sự tác động của quán tính, nghĩa là tốc độ trượt từ bước trước đó. Nói cách khác, chúng ta không chỉ sử dụng $g_k = -\nabla f(x_k)$ mà cả Δ_{k-1}, với một trọng số γ nào đó.

Phương pháp Giảm gradient có quán tính

Như vậy, trong thuật toán gradient có quán tính ta dùng công thức $x_{k+1} := x_k + \Delta_k$, trong đó

$$\Delta_0 := -\alpha \nabla f(x_0)$$
$$\Delta_k := \gamma \Delta_{k-1} - \alpha \nabla f(x_k), \text{ nếu } k \geq 1$$

Thuật toán

Bước 0: Chọn $\epsilon > 0$, $\alpha > 0$,$\gamma > 0$, $x_0 \in \mathbb{R}^n$ và đặt $k := 0$
Bước 1: Nếu $\|g_k\| < \epsilon$ dừng; ngược lại sang Bước 2.
Bước 2: Nếu $k = 0$ đặt $\Delta_0 := -\alpha g_0$. Ngược lại $\Delta_k := \gamma \Delta_{k-1} - \alpha g_k$.
Bước 3: Đặt $x_{k+1} = x_k + \Delta_k$
Bước 4: Đặt $k := k + 1$, sang Bước 1.

Áp dụng thuật toán với $\gamma = 0.9$, $\alpha = 0.1$, $x_0 = 5$, $\epsilon = 0.1$

k	x_k	g_k	Δ_k	k	x_k	g_k	Δ_k
0	5	12.836	-1.283	25	-1.750	-5.286	-0.258
1	3.716	-0.961	-1.059	26	-2.008	-8.254	0.593
2	2.657	-3.535	-.600	27	-1.415	-1.281	0.662
3	2.057	-0.561	-0.484	28	-0.753	5.787	0.017
4	1.574	3.119	-.747	29	-0.736	5.936	-0.578
...	...	...	...	30	-1.315	-0.097	-0.511

Nhận xét

Như vậy ta thấy Thuật toán giảm gradient có quán tính sẽ giúp hòn bi vượt qua được dốc cực tiểu địa phương. Tuy nhiên, có một hạn chế đó là tới gần đích, thuật toán vẫn mất khá nhiều thời gian trước khi dừng lại. Lý do lại cũng chính là vì có quán tính, nên nó dao động qua lại.

Ý tưởng

Trong khi các phương pháp trên đây chủ yếu sử dụng các đạo hàm bậc nhất của f, thì Phương pháp Newton sử dụng cả đạo hàm bậc hai. Ý tưởng của phương pháp là tại điểm x_k ta khai triển Taylor đến bậc 2 để xấp xỉ f bởi một hàm toàn phương q. Sau đó chọn điểm lặp kế tiếp x_{k+1} chính là điểm cực tiểu của q.

Khai triển Taylor

$$f(x) \approx q(x) := f(x_k) + \langle \nabla f(x_k), x - x_k \rangle + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k).$$

$$\nabla q(x) = \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k); \; \nabla^2 q(x) = \nabla^2 f(x_k).$$

Vì vậy nếu $\nabla^2 f(x_k)$ xác định dương thì q là một hàm toàn phương lồi và điểm cực tiểu x_{k+1} của q chính là nghiệm của phương trình

$$\nabla q(x) = 0 \Leftrightarrow \nabla f(x_k) + \nabla^2 f(x_k)(x - x_k) = 0.$$

$$x_{k+1} := x_k - \left(\nabla^2 f(x_k)\right)^{-1} \nabla f(x_k).$$

Kí hiệu

Tại mỗi điểm x_k, ta kí hiệu $g_k := \nabla f(x_k)$; $G_k := \nabla^2 f(x_k)$.

Thuật toán Newton

Bước 0: Chọn $\epsilon > 0$, $x_0 \in \mathbb{R}^n$, đặt $k := 0$ và tính

$$g_0 := \nabla f(x_0); \quad G_0 := \nabla^2 f(x_0).$$

Bước 1: Nếu $\|g_k\| < \epsilon$ dừng; ngược lại sang Bước 2.

Bước 2: Giải phương trình $G_k s_k = -g_k$ để xác định s_k

Bước 3: Tính $x_{k+1} := x_k + s_k$, $g_{k+1} := \nabla f(x_{k+1})$, $G_{k+1} := \nabla^2 f(x_{k+1})$.

Bước 4: Đặt $k := k + 1$, sang Bước 1.

Nhận xét

Nếu f là hàm toàn phương lồi thì $q \equiv f$, nên chỉ cần một bước ta có ngay x_1 là điểm cực tiểu của f.

Tìm cực tiểu hàm $f(x) = \frac{3}{2}x_1^2 + \frac{1}{2}x_2^2 - x_1x_2 - 2x_1$ bằng pp Newton với $\epsilon = 0,1$ điểm xuất phát $x_0 = (-2,4)$.

$$\nabla f(x) = (3x_1 - x_2 - 2; x_2 - x_1); \quad \nabla^2 f(x) = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}.$$

B0: Đặt $x_0 := (-2,4), \epsilon := 0,1, k := 0$.

Tính $g_0 := \nabla f(x_0) = (-12,6)$, $G_0 := \nabla^2 f(x_0) = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$.

B1: Do $\|g_0\| = \sqrt{180} \geq \epsilon$ nên sang Bước 2.

B2: Giải phương trình $G_0 s = -g_0$, ta nhận được

$$s_0 = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 12 \\ -6 \end{pmatrix} = -\begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 3/2 \end{pmatrix} \begin{pmatrix} -12 \\ 6 \end{pmatrix} = \begin{pmatrix} 3 \\ -3 \end{pmatrix}$$

Ví dụ (tiếp tục)

B3: Tính

$$x_1 := x_0 + s_0 = \begin{pmatrix} -2 \\ 4 \end{pmatrix} + \begin{pmatrix} 3 \\ -3 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}; \quad g_1 := \nabla f(x_1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

B4: Đặt $k := k + 1 = 1$ sang Bước 1.

B1: Do $\|g_1\| = 0 < \epsilon$, thuật toán dừng với nghiệm

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$