Line search method

Hoàng Nam Dũng

Khoa Toán - Cơ - Tin học, Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

Line search method

Tại mỗi bước, từ điểm x_k hiện tại, phương pháp line search tính một hướng tìm kiếm (search direction) p_k rồi quyết định sẽ tiến bao xa theo hướng đó.

Line search method

Tại mỗi bước, từ điểm x_k hiện tại, phương pháp line search tính một hướng tìm kiếm (search direction) p_k rồi quyết định sẽ tiến bao xa theo hướng đó. Công thức lặp để tính điểm tiếp theo được cho bởi

$$x_{k+1} = x_k + \alpha_k p_k$$

trong đó $\alpha_k > 0$ được gọi là *độ dài bước (step length)*.

Line search method

Tại mỗi bước, từ điểm x_k hiện tại, phương pháp line search tính một hướng tìm kiếm (search direction) p_k rồi quyết định sẽ tiến bao xa theo hướng đó. Công thức lặp để tính điểm tiếp theo được cho bởi

$$x_{k+1} = x_k + \alpha_k p_k$$

trong đó $\alpha_k > 0$ được gọi là *độ dài bước (step length)*.

Hiệu quả của phương pháp phụ thuộc vào việc chọn hướng p_k và độ dài bước α_k thích hợp.

Line search method

Tại mỗi bước, từ điểm x_k hiện tại, phương pháp line search tính một hướng tìm kiếm (search direction) p_k rồi quyết định sẽ tiến bao xa theo hướng đó. Công thức lặp để tính điểm tiếp theo được cho bởi

$$x_{k+1} = x_k + \alpha_k p_k$$

trong đó $\alpha_k > 0$ được gọi là *độ dài bước (step length)*.

Hiệu quả của phương pháp phụ thuộc vào việc chọn hướng p_k và độ dài bước α_k thích hợp.

Hầu hết các phương pháp line search đòi hỏi p_k là một *hướng giảm (descent direction)*

$$p_k^T \nabla f(x_k) < 0$$

bởi nó sẽ đảm bảo là giá trị hàm f có thể giảm xuống theo hướng này.

Hướng giảm (descent direction)

Giả sử p là một hướng giảm, tức là

$$p^T \nabla f(x_k) < 0.$$

Hướng giảm (descent direction)

Giả sử p là một hướng giảm, tức là

$$p^T \nabla f(x_k) < 0.$$

Theo công thức khai triển Taylor ta có

$$f(x_k + \alpha p) = f(x_k) + \alpha p^T \nabla f(x_k) + O(\alpha^2).$$

Hướng giảm (descent direction)

Giả sử p là một hướng giảm, tức là

$$p^T \nabla f(x_k) < 0.$$

Theo công thức khai triển Taylor ta có

$$f(x_k + \alpha p) = f(x_k) + \alpha p^T \nabla f(x_k) + O(\alpha^2).$$

Suy ra $f(x_k + \alpha p) < f(x_k)$ với $\alpha > 0$ đủ nhỏ. Tức là ta có thể giảm giá trị hàm số f theo hướng p.

Hướng giảm nhanh/dốc nhất (steepest descent direction)

Hướng giảm nhanh nhất là nghiệm của bài toán tối ưu

$$\min_p p^T \nabla f(x_k), \text{ s.t. } \|p\| = 1.$$

Hướng giảm nhanh/dốc nhất (steepest descent direction)

Hướng giảm nhanh nhất là nghiệm của bài toán tối ưu

$$\min_p p^T \nabla f(x_k), \text{ s.t. } \|p\| = 1.$$

Gọi θ là góc giữa p và $\nabla f(x_k)$.

Hướng giảm nhanh/dốc nhất (steepest descent direction)

Hướng giảm nhanh nhất là nghiệm của bài toán tối ưu

$$\min_p p^T \nabla f(x_k), \text{ s.t. } \|p\| = 1.$$

Gọi θ là góc giữa p và $\nabla f(x_k)$. Ta có

$$p^T \nabla f(x_k) = \|p\| \|\nabla f(x_k)\| \cos\theta = \|\nabla f(x_k)\| \cos\theta,$$

Hướng giảm nhanh/dốc nhất (steepest descent direction)

Hướng giảm nhanh nhất là nghiệm của bài toán tối ưu

$$\min_p p^T \nabla f(x_k), \text{ s.t. } \|p\| = 1.$$

Gọi θ là góc giữa p và $\nabla f(x_k)$. Ta có

$$p^T \nabla f(x_k) = \|p\| \|\nabla f(x_k)\| \cos\theta = \|\nabla f(x_k)\| \cos\theta,$$

tức là nó đạt giá trị bé nhất khi $\cos\theta = -1$ hay

$$p = -\frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}.$$

Hướng giảm nhanh/dốc nhất (steepest descent direction)

Hướng giảm nhanh nhất là nghiệm của bài toán tối ưu

$$\min_p p^T \nabla f(x_k), \text{ s.t. } \|p\| = 1.$$

Gọi θ là góc giữa p và $\nabla f(x_k)$. Ta có

$$p^T \nabla f(x_k) = \|p\| \|\nabla f(x_k)\| \cos\theta = \|\nabla f(x_k)\| \cos\theta,$$

tức là nó đạt giá trị bé nhất khi $\cos\theta = -1$ hay

$$p = -\frac{\nabla f(x_k)}{\|\nabla f(x_k)\|}.$$

Phương pháp line search với $p_k = -\nabla f(x_k)$ được gọi là *steepest descent method* hay *gradient descent method*.

Hướng giảm phổ biến

Ta thường chọn hướng giảm có dạng

$$p_k = -B_k^1 \nabla f(x_k)$$

trong đó B_k là một ma trận đối xứng không suy biến.

Hướng giảm phổ biến

Ta thường chọn hướng giảm có dạng

$$p_k = -B_k^1 \nabla f(x_k)$$

trong đó B_k là một ma trận đối xứng không suy biến.

- Phương pháp gradient descent: $B_k = I$
- Phương pháp Newton: $B_k = \nabla^2 f(x_k)$
- Phương pháp tựa Newton: B_k xấp xỉ ma trận Hessian $\nabla^2 f(x_k)$.

Hướng giảm phổ biến

Ta thường chọn hướng giảm có dạng

$$p_k = -B_k^1 \nabla f(x_k)$$

trong đó B_k là một ma trận đối xứng không suy biến.

- Phương pháp gradient descent: $B_k = I$
- Phương pháp Newton: $B_k = \nabla^2 f(x_k)$
- Phương pháp tựa Newton: B_k xấp xỉ ma trận Hessian $\nabla^2 f(x_k)$.

Phần tiếp theo chúng ta sẽ tìm hiểu xem chọn hướng và độ dài bước như thế nào để thuật toán hội tụ.

Lựa chọn độ dài bước

Ta phải cân nhắc lựa chọn giữa hai yếu tố

- Tìm α_k để cải thiện đáng kể giá trị hàm mục tiêu
- Đồng thời không được mất quá nhiều thời gian.

Lựa chọn độ dài bước

Ta phải cân nhắc lựa chọn giữa hai yếu tố

- Tìm α_k để cải thiện đáng kể giá trị hàm mục tiêu
- Đồng thời không được mất quá nhiều thời gian.

Lý tưởng thì α_k là nghiệm của bài toán tối ưu

$$\min \phi(\alpha) = f(x_k + \alpha p_k), \ \alpha > 0.$$

Lựa chọn độ dài bước

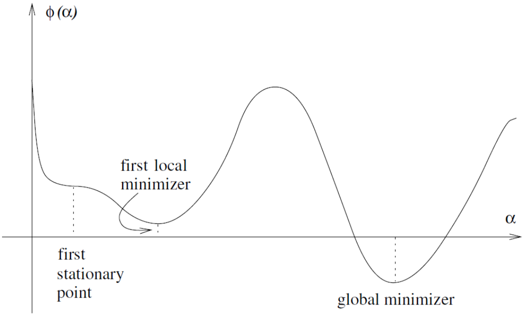
Ta phải cân nhắc lựa chọn giữa hai yếu tố

- Tìm α_k để cải thiện đáng kể giá trị hàm mục tiêu
- Đồng thời không được mất quá nhiều thời gian.

Lý tưởng thì α_k là nghiệm của bài toán tối ưu

$$\min \phi(\alpha) = f(x_k + \alpha p_k), \ \alpha > 0.$$

Tuy nhiên nó quá tốn kém (ví dụ xem hình vẽ).



Lựa chọn độ dài bước

Để tìm cực tiểu địa phương với độ chính xác vừa phải có thể ta phải tính rất nhiều lần hàm mục tiêu f và gradient ∇f.

Lựa chọn độ dài bước

Để tìm cực tiểu địa phương với độ chính xác vừa phải có thể ta phải tính rất nhiều lần hàm mục tiêu f và gradient ∇f.

Chiến lược thực tế hơn đó là chúng ta tìm các độ dài bước đủ tốt theo nghĩa hàm mục tiêu phải giảm đủ nhiều.

Lựa chọn độ dài bước

Để tìm cực tiểu địa phương với độ chính xác vừa phải có thể ta phải tính rất nhiều lần hàm mục tiêu f và gradient ∇f.

Chiến lược thực tế hơn đó là chúng ta tìm các độ dài bước đủ tốt theo nghĩa hàm mục tiêu phải giảm đủ nhiều.

Yêu cầu giảm $f(x_k + \alpha_k p_k) < f(x_k)$ là chưa đủ.

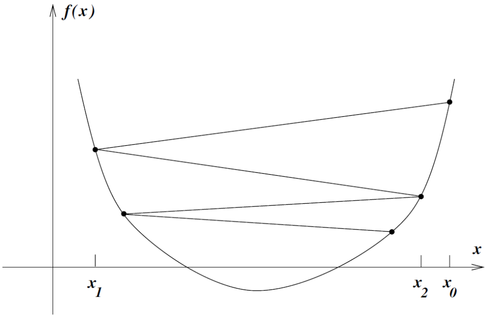
Lựa chọn độ dài bước

Để tìm cực tiểu địa phương với độ chính xác vừa phải có thể ta phải tính rất nhiều lần hàm mục tiêu f và gradient ∇f.

Chiến lược thực tế hơn đó là chúng ta tìm các độ dài bước đủ tốt theo nghĩa hàm mục tiêu phải giảm đủ nhiều.

Yêu cầu giảm $f(x_k + \alpha_k p_k) < f(x_k)$ là chưa đủ.

Ví dụ:
Giá trị tối ưu $f^* = -1$ nhưng một chuỗi x_k với $f(x_k) = \frac{5}{k}$ chỉ hội tụ đến 0.



Các điều kiện Wolfe

Điều kiện 1: hàm mục tiêu phải giảm đủ nhiều

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$

với $c_1 \in (0, 1)$.

Các điều kiện Wolfe

Điều kiện 1: hàm mục tiêu phải giảm đủ nhiều

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$

với $c_1 \in (0, 1)$.

Lưu ý rằng p_k là hướng giảm, $\nabla f_k^T p_k < 0$.

Các điều kiện Wolfe

Điều kiện 1: hàm mục tiêu phải giảm đủ nhiều

$$f(x_k + \alpha_k p_k) \le f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$

với $c_1 \in (0, 1)$.

Lưu ý rằng p_k là hướng giảm, $\nabla f_k^T p_k < 0$.

Độ giảm của f tỉ lệ với độ dài bước α_k và $\nabla f_k^T p_k$.

Các điều kiện Wolfe

Điều kiện 1: hàm mục tiêu phải giảm đủ nhiều

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$

với $c_1 \in (0, 1)$.

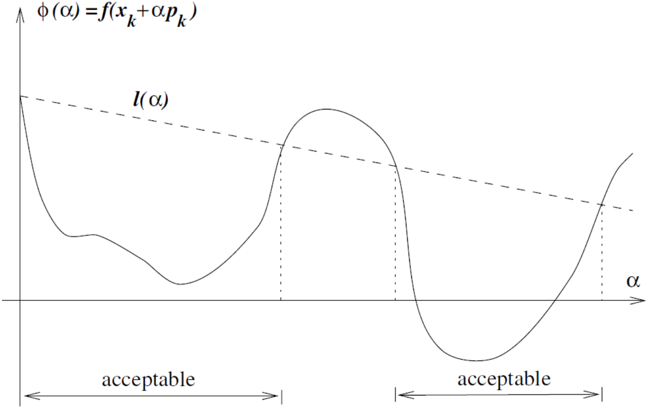
Lưu ý rằng p_k là hướng giảm, $\nabla f_k^T p_k < 0$.

Độ giảm của f tỉ lệ với độ dài bước α_k và $\nabla f_k^T p_k$.

Trong thực tế c_1 được chọn khá nhỏ, ví dụ $c_1 = 10^{-4}$.

Điều kiện Wolfe - minh họa điều kiện 1

$l(\alpha) := f(x_k) + c_1\alpha\nabla f_k^T p_k$ (là một đường thẳng có hệ số góc âm).



Hình 1: Điều kiện giảm đủ

Điều kiện Wolfe

Hình 1 cho thấy Điều kiện 1 thỏa mãn với α đủ nhỏ.

Điều kiện Wolfe

Hình 1 cho thấy Điều kiện 1 thỏa mãn với α đủ nhỏ.

Để loại trừ những bước quá ngắn ta đòi hỏi thêm Điều kiện 2 (curvature condition)

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k$$

với $c_2 \in (c_1, 1)$.

Điều kiện Wolfe

Hình 1 cho thấy Điều kiện 1 thỏa mãn với α đủ nhỏ.

Để loại trừ những bước quá ngắn ta đòi hỏi thêm Điều kiện 2 (curvature condition)

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k$$

với $c_2 \in (c_1, 1)$.

Lưu ý rằng vế trái chính là $\phi'(\alpha_k)$.

Điều kiện Wolfe

Hình 1 cho thấy Điều kiện 1 thỏa mãn với α đủ nhỏ.

Để loại trừ những bước quá ngắn ta đòi hỏi thêm Điều kiện 2 (curvature condition)

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k$$

với $c_2 \in (c_1, 1)$.

Lưu ý rằng vế trái chính là $\phi'(\alpha_k)$. Điều kiện này đòi hỏi độ dốc của ϕ tại α_k lớn hơn hoặc bằng c_2 lần độ dốc ban đầu $\phi'(0)$.

Điều kiện Wolfe

Hình 1 cho thấy Điều kiện 1 thỏa mãn với α đủ nhỏ.

Để loại trừ những bước quá ngắn ta đòi hỏi thêm Điều kiện 2 (curvature condition)

$$\nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k$$

với $c_2 \in (c_1, 1)$.

Lưu ý rằng vế trái chính là $\phi'(\alpha_k)$. Điều kiện này đòi hỏi độ dốc của ϕ tại α_k lớn hơn hoặc bằng c_2 lần độ dốc ban đầu $\phi'(0)$.

Trong thực tế ta chọn

- $c_2 = 0.9$ với thuật toán Newton, tựa Newton
- $c_2 = 0.1$ với thuật toán conjugate gradient.

Diễn giải điều kiện 2

$$\phi'(\alpha_k) = \nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k = c_2 \phi'(0)$$

Lưu ý rằng $\phi'(0) < 0$ và $0 < c_2 < 1$.

Diễn giải điều kiện 2

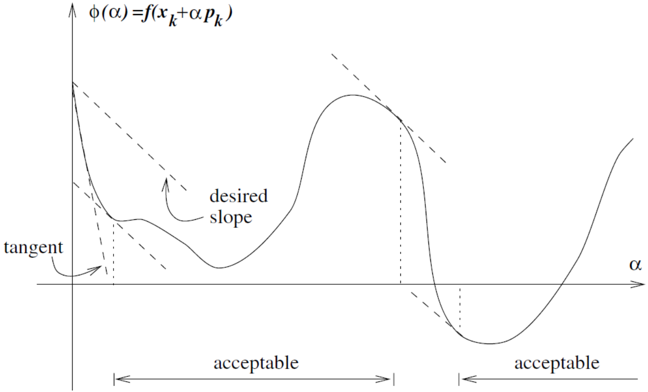
$$\phi'(\alpha_k) = \nabla f(x_k + \alpha_k p_k)^T p_k \geq c_2 \nabla f_k^T p_k = c_2 \phi'(0)$$

Lưu ý rằng $\phi'(0) < 0$ và $0 < c_2 < 1$.

Như vậy để thỏa mãn điều kiện thì α_k sẽ không quá gần 0.

Đồng thời điều kiện đó cũng loại α có độ dốc đủ lớn $\phi'(\alpha)$. Điều này hợp lí vì đó là dấu hiệu ta có thể tiếp tục đi xa hơn theo hướng p_k.

Minh họa điều kiện 2



Hình 2: Độ dài bước thỏa mãn điều kiện 2

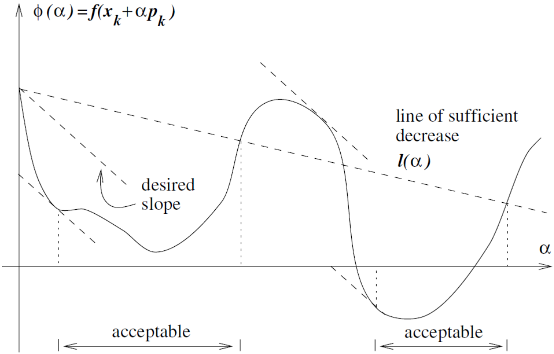
Minh họa điều kiện Wolfe

Điều kiện Wolfe

$$f(x_k + \alpha_k p_k) \le f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$

$$\nabla f(x_k + \alpha_k p_k)^T p_k \ge c_2 \nabla f_k^T p_k$$

với $0 < c_1 < c_2 < 1$.



Hình 3: Độ dài bước thỏa mãn hai điều kiện Wolfe

Điều kiện Wolfe mạnh

Hình 3 cho thấy có những độ dài bước thỏa mãn điều kiện Wolfe nhưng nó không đủ gần một cực tiểu.

Hình 3 cho thấy có những độ dài bước thỏa mãn điều kiện Wolfe nhưng nó không đủ gần một cực tiểu.

Ta có thể điều chỉnh điều kiện 2 để ép α_k nằm trong lân cận của một cực tiểu địa phương.

Điều kiện Wolfe mạnh

Hình 3 cho thấy có những độ dài bước thỏa mãn điều kiện Wolfe nhưng nó không đủ gần một cực tiểu.

Ta có thể điều chỉnh điều kiện 2 để ép α_k nằm trong lân cận của một cực tiểu địa phương.

Điều kiện Wolfe mạnh:

$$f(x_k + \alpha_k p_k) \leq f(x_k) + c_1 \alpha_k \nabla f_k^T p_k,$$
$$|\nabla f(x_k + \alpha_k p_k)^T p_k| \leq c_2 \nabla |f_k^T p_k|$$

với $0 < c_1 < c_2 < 1$.

Sự tồn tại độ dài bước thỏa mãn điều kiện Wolfe

Bổ đề

Cho $f : \mathbb{R}^n \to \mathbb{R}$ *khả vi liên tục,* p_k *là một hướng giảm tại* x_k *và giả sử rằng* f *bị chặn dưới dọc theo tia* $\{x_k + \alpha_k \,|\, \alpha > 0\}$*. Nếu* $0 < c_1 < c_2 < 1$ *thì tồn tại độ dài bước thỏa mãn điều kiện Wolfe và điều kiện Wolfe mạnh.*

Chứng minh.

Xem chứng minh Bổ đề 3.1, sách Nocedal. □

Điều kiện Goldstein

Với $0 < c < \frac{1}{2}$, điều kiện Goldstein đòi hỏi α_k thỏa mãn

$$f(x_k) + (1-c)\alpha_k \nabla f_k^T p_k \leq f(x_k + \alpha_k p_k) \leq f(x_k) + c\alpha_k \nabla f_k^T p_k.$$

Về ý nghĩa và nhận xét xem trang 36 sách của Nocedal.

Backtracking line search

Phương pháp backtracking tìm độ dài bước thỏa mãn điều kiện Wolfe thứ nhất theo một cách thích hợp (mà không cần đòi hỏi điều kiện 2).

Backtracking line search

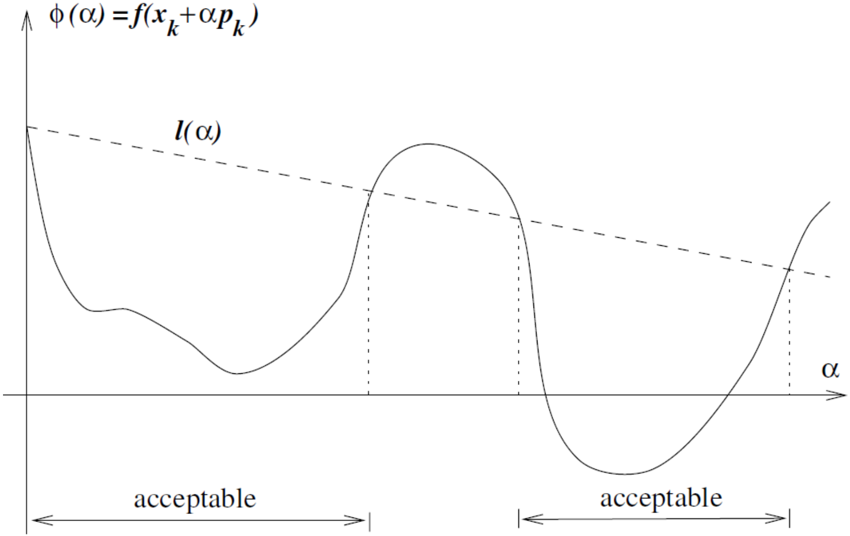
Phương pháp backtracking tìm độ dài bước thỏa mãn điều kiện Wolfe thứ nhất theo một cách thích hợp (mà không cần đòi hỏi điều kiện 2).

Algorithm 2 Backtracking line search

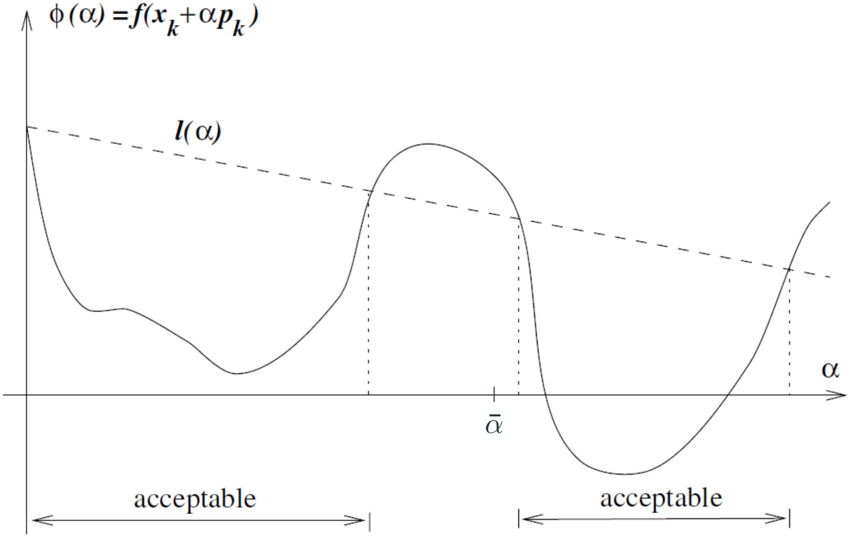
Chọn $\bar{\alpha} > 0$, $\rho \in (0, 1)$, $c \in (0, 1)$
Lấy $\alpha := \bar{\alpha}$
while $f(x_k + \alpha_k p_k) > f(x_k) + c\alpha_k \nabla f_k^T p_k$ **do**
    $\alpha := \rho\alpha$
end while
$\alpha_k := \alpha$

Nhận xét: Sau một số bước thì α sẽ đủ nhỏ để thỏa mãn điều kiện.

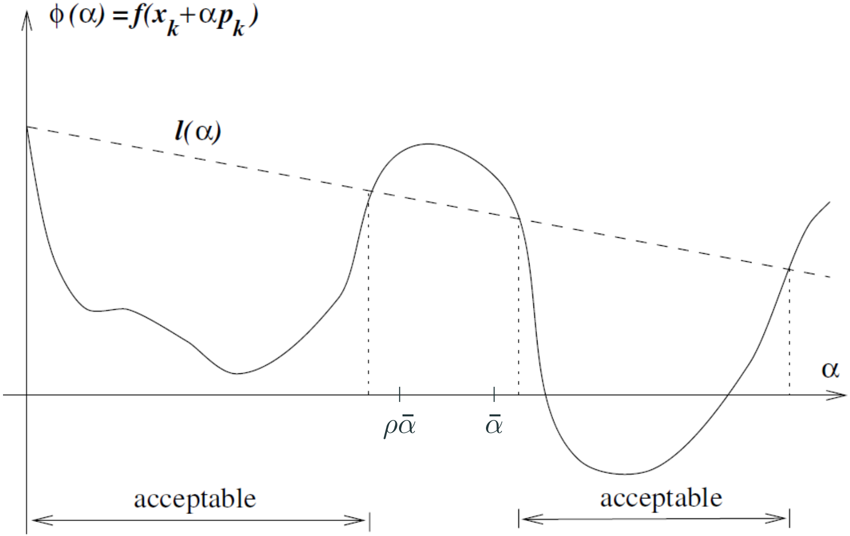
Minh hoạ backtracking line search



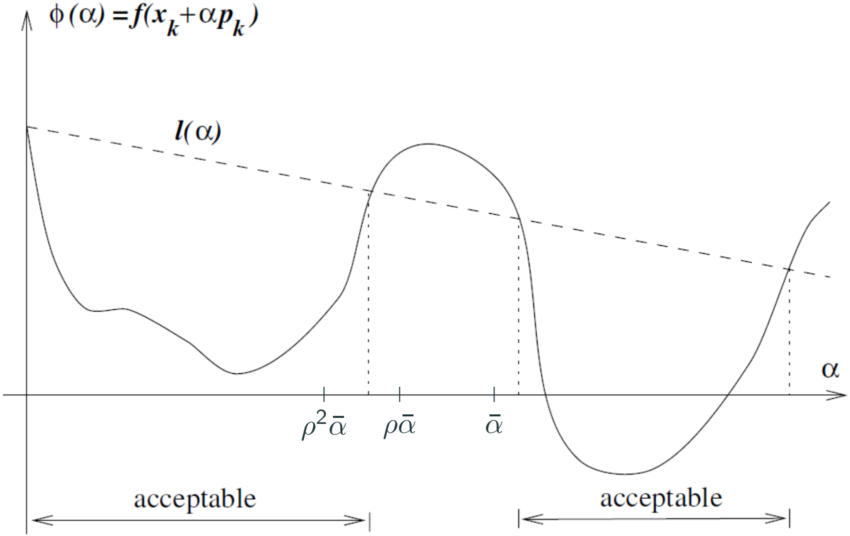
Minh hoạ backtracking line search



Minh hoạ backtracking line search



Minh hoạ backtracking line search



Chọn tham số cho backtracking line search

Chọn $\bar{\alpha}$:

- $\bar{\alpha} = 1$ với phương pháp Newton hay tựa Newton
- Có các giá trị khác nhau với thuật toán gradient.

Chọn tham số cho backtracking line search

Chọn $\bar{\alpha}$:

- $\bar{\alpha} = 1$ với phương pháp Newton hay tựa Newton
- Có các giá trị khác nhau với thuật toán gradient.

ρ có thể chọn khác nhau trong mỗi bước lặp của phương pháp line search.

Các phương pháp chọn độ dài bước

Đọc thêm mục 3.5 sách của Nocedal (quan trọng nhưng không đủ thời gian để trình bày trên lớp).

Sự hội tụ của phương pháp line search và tốc độ hội tụ

Mục 3.2 sách của Nocedal chứng minh sự hội tụ của phương pháp line search khi độ dài bước thỏa mãn điều kiện Wolfe.

Sự hội tụ của phương pháp line search và tốc độ hội tụ

Mục 3.2 sách của Nocedal chứng minh sự hội tụ của phương pháp line search khi độ dài bước thỏa mãn điều kiện Wolfe.

Ta sẽ chứng minh công thức tốc độ hội tụ khi xét từng thuật toán

- Gradient descent
- Newton
- Tựa Newton.

Tài liệu tham khảo

Chương 3, J. Nocedal and S. Wright, Numerical Optimization, Springer.