Bài toán tối ưu không ràng buộc

Hoàng Nam Dũng

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

Bài toán tối ưu không ràng buộc (unconstrained)

$$\min_{x} f(x)$$

với $f : \mathbb{R}^n \to \mathbb{R}$ là một hàm trơn (smooth).

Bài toán tối ưu không ràng buộc (unconstrained)

$$\min_{x} f(x)$$

với $f : \mathbb{R}^n \to \mathbb{R}$ là một hàm trơn (smooth).

Định nghĩa

x^* *được gọi là* cực tiểu toàn cục *nếu*

$$f(x^*) \leq f(x), \ \forall x.$$

Bài toán tối ưu không ràng buộc (unconstrained)

$$\min_x f(x)$$

với $f : \mathbb{R}^n \to \mathbb{R}$ là một hàm trơn (smooth).

Định nghĩa

x^* *được gọi là* cực tiểu toàn cục *nếu*

$$f(x^*) \leq f(x), \ \forall x.$$

x^* *được gọi là* cực tiểu địa phương *nếu tồn tại một lân cận* $\mathcal{N}$ *của* x^* *sao cho*

$$f(x^*) \leq f(x), \ \forall x \in \mathcal{N}.$$

Bài toán tối ưu không ràng buộc (unconstrained)

$$\min_x f(x)$$

với $f : \mathbb{R}^n \to \mathbb{R}$ là một hàm trơn (smooth).

Định nghĩa

x^* được gọi là cực tiểu toàn cục *nếu*

$$f(x^*) \leq f(x), \ \forall x.$$

x^* được gọi là cực tiểu địa phương *nếu tồn tại một lân cận $\mathcal{N}$ của x^* sao cho*

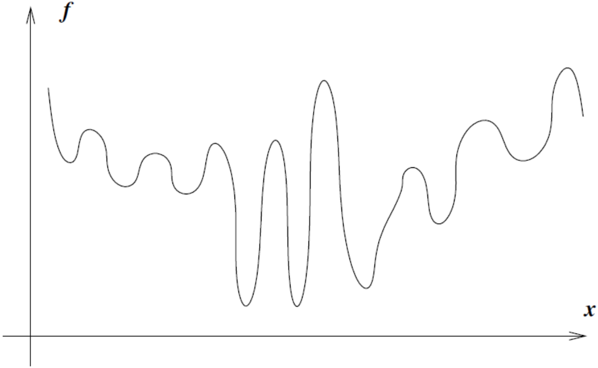
$$f(x^*) \leq f(x), \ \forall x \in \mathcal{N}.$$

x^* được gọi là cực tiểu địa phương mạnh (hay ngặt) *nếu tồn tại một lân cận $\mathcal{N}$ của x^* sao cho*

$$f(x^*) < f(x), \ \forall x \in \mathcal{N} \backslash \{x^*\}.$$

Ví dụ

Hàm số dưới đây có nhiều cực tiểu địa phương và khó để tìm cực tiểu toàn cục.



Điều kiện cực tiểu địa phương

Định lý (Khai triển Taylor)

Cho $f : \mathbb{R}^n \to \mathbb{R}$ *khả vi liên tục và* $p \in \mathbb{R}^n$. *Ta có*

$$f(x+p) = f(x) + \nabla f(x+tp)^T p,$$

với $t \in (0,1)$ *nào đó.*

Định lý (Khai triển Taylor)

Cho $f : \mathbb{R}^n \to \mathbb{R}$ *khả vi liên tục và* $p \in \mathbb{R}^n$. *Ta có*

$$f(x+p) = f(x) + \nabla f(x+tp)^T p,$$

với $t \in (0,1)$ *nào đó. Nếu* f *khả vi liên tục hai lần thì*

$$f(x+p) = f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x+tp) p,$$

với $t \in (0,1)$ *nào đó.*

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc nhất)

Nếu x^* *là một cực tiểu địa phương và* f *khả vi liên tục trong một lân cận mở của* x^* *thì* $\nabla f(x^*) = 0$.

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc nhất)

Nếu x^* *là một cực tiểu địa phương và* f *khả vi liên tục trong một lân cận mở của* x^* *thì* $\nabla f(x^*) = 0$.

Chứng minh.

Giả sử $\nabla f(x^*) \neq 0$. Chọn $p = -\nabla f(x^*)$ thì

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0.$$

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc nhất)

Nếu x^* là một cực tiểu địa phương và f khả vi liên tục trong một lân cận mở của x^* thì $\nabla f(x^*) = 0$.

Chứng minh.

Giả sử $\nabla f(x^*) \neq 0$. Chọn $p = -\nabla f(x^*)$ thì

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0.$$

Từ đó và do ∇f liên tục trong lân cận của x^*, tồn tại $T > 0$ sao cho

$$p^T \nabla f(x^* + sp) < 0, \ \forall s \in [0, T].$$

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc nhất)

Nếu x^* là một cực tiểu địa phương và f khả vi liên tục trong một lân cận mở của x^* thì $\nabla f(x^*) = 0$.

Chứng minh.

Giả sử $\nabla f(x^*) \neq 0$. Chọn $p = -\nabla f(x^*)$ thì

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0.$$

Từ đó và do ∇f liên tục trong lân cận của x^*, tồn tại $T > 0$ sao cho

$$p^T \nabla f(x^* + sp) < 0, \ \forall s \in [0, T].$$

Với mỗi $t \in (0, T]$ theo định lý Taylor tồn tại $s \in (0, t)$ sao cho

$$f(x^* + tp) = f(x^*) + t\nabla f(x^* + sp)^T p$$

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc nhất)

Nếu x^* là một cực tiểu địa phương và f khả vi liên tục trong một lân cận mở của x^* thì $\nabla f(x^*) = 0$.

Chứng minh.

Giả sử $\nabla f(x^*) \neq 0$. Chọn $p = -\nabla f(x^*)$ thì

$$p^T \nabla f(x^*) = -\|\nabla f(x^*)\|^2 < 0.$$

Từ đó và do ∇f liên tục trong lân cận của x^*, tồn tại $T > 0$ sao cho

$$p^T \nabla f(x^* + sp) < 0, \ \forall s \in [0, T].$$

Với mỗi $t \in (0, T]$ theo định lý Taylor tồn tại $s \in (0, t)$ sao cho

$$f(x^* + tp) = f(x^*) + t\nabla f(x^* + sp)^T p < f(x^*),$$

mâu thuẫn. $\square$

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc hai)

Nếu x^* *là một cực tiểu địa phương và* $\nabla^2 f$ *tồn tại và liên tục trong một lân cận mở của* x^* *thì* $\nabla f(x^*) = 0$ *và* $\nabla^2 f(x^*)$ *là nửa xác định dương (positive semidefinite).*

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc hai)

Nếu x^* là một cực tiểu địa phương và $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^* thì $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là nửa xác định dương (positive semidefinite).

Chứng minh.

Từ định lý trước ta có $\nabla f(x^*) = 0$.

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc hai)

Nếu x^* là một cực tiểu địa phương và $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^* thì $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là nửa xác định dương (positive semidefinite).

Chứng minh.

Từ định lý trước ta có $\nabla f(x^*) = 0$. Giả sử $\nabla^2 f(x^*)$ không phải nửa xác định dương, tức là tồn tại p sao cho $p^T \nabla^2 f(x^*) p < 0$.

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc hai)

Nếu x^* là một cực tiểu địa phương và $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^* thì $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là nửa xác định dương (positive semidefinite).

Chứng minh.

Từ định lý trước ta có $\nabla f(x^*) = 0$. Giả sử $\nabla^2 f(x^*)$ không phải nửa xác định dương, tức là tồn tại p sao cho $p^T \nabla^2 f(x^*) p < 0$. Vì $\nabla^2 f$ liên tục quanh x^* nên tồn tại $T > 0$ sao cho

$$p^T \nabla^2 f(x^* + sp) p < 0, \ \forall s \in [0, T].$$

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc hai)

Nếu x^* là một cực tiểu địa phương và $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^* thì $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là nửa xác định dương (positive semidefinite).

Chứng minh.

Từ định lý trước ta có $\nabla f(x^*) = 0$. Giả sử $\nabla^2 f(x^*)$ không phải nửa xác định dương, tức là tồn tại p sao cho $p^T \nabla^2 f(x^*) p < 0$. Vì $\nabla^2 f$ liên tục quanh x^* nên tồn tại $T > 0$ sao cho

$$p^T \nabla^2 f(x^* + sp) p < 0, \ \forall s \in [0, T].$$

Áp dụng định lý Taylor ta có với mỗi $t \in (0, T]$ tồn tại $s \in (0, t)$ sao cho

$$f(x^* + tp) = f(x^*) + t \nabla f(x^*)^T p + \frac{1}{2} t^2 p^T \nabla^2 f(x^* + sp) p$$

Điều kiện cực tiểu địa phương

Định lý (Điều kiện cần bậc hai)

Nếu x^* là một cực tiểu địa phương và $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^* thì $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là nửa xác định dương (positive semidefinite).

Chứng minh.

Từ định lý trước ta có $\nabla f(x^*) = 0$. Giả sử $\nabla^2 f(x^*)$ không phải nửa xác định dương, tức là tồn tại p sao cho $p^T \nabla^2 f(x^*) p < 0$. Vì $\nabla^2 f$ liên tục quanh x^* nên tồn tại $T > 0$ sao cho

$$p^T \nabla^2 f(x^* + sp) p < 0, \ \forall s \in [0, T].$$

Áp dụng định lý Taylor ta có với mỗi $t \in (0, T]$ tồn tại $s \in (0, t)$ sao cho

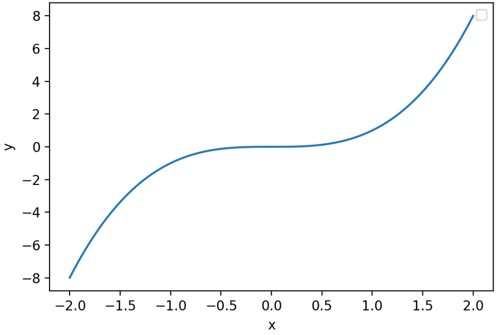
$$f(x^* + tp) = f(x^*) + t\nabla f(x^*)^T p + \frac{1}{2} t^2 p^T \nabla^2 f(x^* + sp) p < f(x^*),$$

mâu thuẫn. □

Điều kiện cực tiểu địa phương

Điều kiện cần bậc 2 không phải là điều kiện đủ.

Ví dụ với $f(x) = x^3$ thì $f'(0) = 0$ và $f''(0) = 0$, tức là $x = 0$ thỏa mãn điều kiện cần bậc 2, tuy nhiên nó không phải là một điểm cực tiểu địa phương của x^3.



Điều kiện cực tiểu địa phương

Định lý (Điều kiện đủ bậc hai)

Nếu $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^*, $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là xác định dương (positive definite) thì x^* là một cực tiểu địa phương ngặt.

Điều kiện cực tiểu địa phương

Định lý (Điều kiện đủ bậc hai)

Nếu $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^*, $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là xác định dương (positive definite) thì x^* là một cực tiểu địa phương ngặt.

Chứng minh.

Vì ma trận Hessian $\nabla^2 f$ liên tục và xác định dương tại x^* nên ta có thể chọn $r > 0$ sao cho $\nabla^2 f(x)$ cũng xác định dương với mọi x thuộc hình cầu $B := \{y \in \mathbb{R}^n \mid \|y - x^*\| \leq r\}$.

Điều kiện cực tiểu địa phương

Định lý (Điều kiện đủ bậc hai)

Nếu $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^*, $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là xác định dương (positive definite) thì x^* là một cực tiểu địa phương ngặt.

Chứng minh.

Vì ma trận Hessian $\nabla^2 f$ liên tục và xác định dương tại x^* nên ta có thể chọn $r > 0$ sao cho $\nabla^2 f(x)$ cũng xác định dương với mọi x thuộc hình cầu $B := \{y \in \mathbb{R}^n \mid \|y - x^*\| \leq r\}$.

Với $p \in \mathbb{R}^n$ thỏa mãn $p \neq 0$ và $\|p\| < r$, ta có $x^* + tp \in B$ với mọi $t \in (0, 1)$.

Điều kiện cực tiểu địa phương

Định lý (Điều kiện đủ bậc hai)

Nếu $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^*, $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là xác định dương (positive definite) thì x^* là một cực tiểu địa phương ngặt.

Chứng minh.

Vì ma trận Hessian $\nabla^2 f$ liên tục và xác định dương tại x^* nên ta có thể chọn $r > 0$ sao cho $\nabla^2 f(x)$ cũng xác định dương với mọi x thuộc hình cầu $B := \{y \in \mathbb{R}^n \mid \|y - x^*\| \leq r\}$.
Với $p \in \mathbb{R}^n$ thỏa mãn $p \neq 0$ và $\|p\| < r$, ta có $x^* + tp \in B$ với mọi $t \in (0, 1)$. Do đó

$$p^T \nabla^2 f(x^* + tp) p > 0, \ \forall t \in (0, 1).$$

Điều kiện cực tiểu địa phương

Định lý (Điều kiện đủ bậc hai)

Nếu $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^*, $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là xác định dương (positive definite) thì x^* là một cực tiểu địa phương ngặt.

Chứng minh.

Vì ma trận Hessian $\nabla^2 f$ liên tục và xác định dương tại x^* nên ta có thể chọn $r > 0$ sao cho $\nabla^2 f(x)$ cũng xác định dương với mọi x thuộc hình cầu $B := \{y \in \mathbb{R}^n \mid \|y - x^*\| \leq r\}$.

Với $p \in \mathbb{R}^n$ thỏa mãn $p \neq 0$ và $\|p\| < r$, ta có $x^* + tp \in B$ với mọi $t \in (0, 1)$. Do đó

$$p^T \nabla^2 f(x^* + tp) p > 0, \ \forall t \in (0, 1).$$

Theo định lý Taylor tồn tại $t \in (0, 1)$ sao cho

$$f(x^* + p) = f(x^*) + \nabla f(x^*)^T p + \frac{1}{2} p^T \nabla^2 f(x^* + tp) p$$

Điều kiện cực tiểu địa phương

Định lý (Điều kiện đủ bậc hai)

Nếu $\nabla^2 f$ tồn tại và liên tục trong một lân cận mở của x^*, $\nabla f(x^*) = 0$ và $\nabla^2 f(x^*)$ là xác định dương (positive definite) thì x^* là một cực tiểu địa phương ngặt.

Chứng minh.

Vì ma trận Hessian $\nabla^2 f$ liên tục và xác định dương tại x^* nên ta có thể chọn $r > 0$ sao cho $\nabla^2 f(x)$ cũng xác định dương với mọi x thuộc hình cầu $B := \{y \in \mathbb{R}^n \mid \|y - x^*\| \le r\}$.

Với $p \in \mathbb{R}^n$ thỏa mãn $p \neq 0$ và $\|p\| < r$, ta có $x^* + tp \in B$ với mọi $t \in (0,1)$. Do đó

$$p^T \nabla^2 f(x^* + tp) p > 0, \ \forall t \in (0,1).$$

Theo định lý Taylor tồn tại $t \in (0,1)$ sao cho

$$f(x^* + p) = f(x^*) + \nabla f(x^*)^T p + \frac{1}{2} p^T \nabla^2 f(x^* + tp) p > f(x^*).$$

Lưu ý

Điều kiện đủ bậc 2 không phải là điều kiện cần cho cực tiểu địa phương ngặt.

Lưu ý

Điều kiện đủ bậc 2 không phải là điều kiện cần cho cực tiểu địa phương ngặt.

Hàm $f(x) = x^4$ có cực tiểu ngặt tại $x = 0$ tuy nhiên đạo hàm bậc 2 của nó bằng 0, tức là không xác định dương.

Cực tiểu của hàm lồi

Định nghĩa

Cho một hàm f khả vi. Một điểm x thỏa mãn $\nabla f(x) = 0$ gọi là một điểm dừng.

Cực tiểu của hàm lồi

Định nghĩa

Cho một hàm f khả vi. Một điểm x thỏa mãn $\nabla f(x) = 0$ gọi là một điểm dừng.

Định lý

Nếu f lồi thì mọi điểm cực tiểu địa phương của f cũng là cực tiểu toàn cục. Ngoài ra nếu f khả vi thì mỗi điểm dừng là một cực tiểu toàn cục.

Chứng minh.

Xem định lý 2.5 sách của Nocedal. $\square$

Cực tiểu của hàm lồi

Định nghĩa

Cho một hàm f khả vi. Một điểm x thỏa mãn $\nabla f(x) = 0$ gọi là một điểm dừng.

Định lý

Nếu f lồi thì mọi điểm cực tiểu địa phương của f cũng là cực tiểu toàn cục. Ngoài ra nếu f khả vi thì mỗi điểm dừng là một cực tiểu toàn cục.

Chứng minh.

Xem định lý 2.5 sách của Nocedal. $\square$

Câu hỏi: Với cực đại thì kết quả tương tự thế nào?

Tổng quan về thuật toán

- Các thuật toán đều đòi hỏi được cung cấp một điểm xuất phát x_0.

Tổng quan về thuật toán

- Các thuật toán đều đòi hỏi được cung cấp một điểm xuất phát x_0.
- Từ đó thuật toán tạo ra chuỗi $x_1, x_2, \ldots$.

Tổng quan về thuật toán

- Các thuật toán đều đòi hỏi được cung cấp một điểm xuất phát x_0.
- Từ đó thuật toán tạo ra chuỗi $x_1, x_2, \ldots$.
- Thuật toán sẽ dừng lại nếu không có sự cải thiện hay có vẻ nghiệm đã xấp xỉ đủ tốt.

Tổng quan về thuật toán

- Các thuật toán đều đòi hỏi được cung cấp một điểm xuất phát x_0.
- Từ đó thuật toán tạo ra chuỗi $x_1, x_2, \ldots$.
- Thuật toán sẽ dừng lại nếu không có sự cải thiện hay có vẻ nghiệm đã xấp xỉ đủ tốt.
- Có 2 chiến thuật cơ bản để tính điểm tiếp theo x_{k+1} từ điểm hiện tại x_k, đó là *line search* và *trust region*.

Line search strategy

Ở bước k thuật toán chọn một hướng p_k và tìm kiếm theo hướng đó từ giá trị hiện tại x_k để tìm được điểm tiếp theo với hàm mục tiêu thấp hơn.

Line search strategy

Ở bước k thuật toán chọn một hướng p_k và tìm kiếm theo hướng đó từ giá trị hiện tại x_k để tìm được điểm tiếp theo với hàm mục tiêu thấp hơn.

Để tìm khoảng cách đi theo hướng này ta có thể giải bài toán tối ưu **1 chiều**

$$\min_{\alpha \geq 0} f(x_k + \alpha p_k).$$

Line search strategy

Ở bước k thuật toán chọn một hướng p_k và tìm kiếm theo hướng đó từ giá trị hiện tại x_k để tìm được điểm tiếp theo với hàm mục tiêu thấp hơn.

Để tìm khoảng cách đi theo hướng này ta có thể giải bài toán tối ưu **1 chiều**

$$\min_{\alpha \geq 0} f(x_k + \alpha p_k).$$

Thực tế ta sẽ không giải chính xác vì nó tốn kém và không cần thiết. Thay vào đó ta chỉ giải xấp xỉ (tương đối lỏng) qua một số bước lặp.

Trust region strategy

Ở bước k ta xây dựng một hàm xấp xỉ m_k của f. Hàm này xấp xỉ f tốt quanh x_k và có thể tệ khi ở xa.

Trust region strategy

Ở bước k ta xây dựng một hàm xấp xỉ m_k của f. Hàm này xấp xỉ f tốt quanh x_k và có thể tệ khi ở xa. Do đó ta chỉ tìm cực tiểu của m_k trong miền quanh x_k qua việc xét bài toán

$$\min_p m_k(x_k + p), \text{ sao cho } x_k + p \text{ thuộc trust region.}$$

Trust region strategy

Ở bước k ta xây dựng một hàm xấp xỉ m_k của f. Hàm này xấp xỉ f tốt quanh x_k và có thể tệ khi ở xa. Do đó ta chỉ tìm cực tiểu của m_k trong miền quanh x_k qua việc xét bài toán

$$\min_p m_k(x_k + p), \text{ sao cho } x_k + p \text{ thuộc trust region.}$$

Nếu điểm được tạo ra không có giá trị hàm mục tiêu giảm đủ nhiều thì ta kết luận rằng trust region quá lớn và phải thu hẹp lại và giải lại bài toán tối ưu trên.

Trust region strategy

Ở bước k ta xây dựng một hàm xấp xỉ m_k của f. Hàm này xấp xỉ f tốt quanh x_k và có thể tệ khi ở xa. Do đó ta chỉ tìm cực tiểu của m_k trong miền quanh x_k qua việc xét bài toán

$$\min_p m_k(x_k + p), \text{ sao cho } x_k + p \text{ thuộc trust region.}$$

Nếu điểm được tạo ra không có giá trị hàm mục tiêu giảm đủ nhiều thì ta kết luận rằng trust region quá lớn và phải thu hẹp lại và giải lại bài toán tối ưu trên.

Thông thường trust region là một hình cầu định nghĩa bởi $\|p\|_2 \leq \Delta$ với $\Delta > 0$ được gọi là bán kính trust region.

Trust region strategy

Ở bước k ta xây dựng một hàm xấp xỉ m_k của f. Hàm này xấp xỉ f tốt quanh x_k và có thể tệ khi ở xa. Do đó ta chỉ tìm cực tiểu của m_k trong miền quanh x_k qua việc xét bài toán

$$\min_p m_k(x_k + p), \text{ sao cho } x_k + p \text{ thuộc trust region.}$$

Nếu điểm được tạo ra không có giá trị hàm mục tiêu giảm đủ nhiều thì ta kết luận rằng trust region quá lớn và phải thu hẹp lại và giải lại bài toán tối ưu trên.

Thông thường trust region là một hình cầu định nghĩa bởi $\|p\|_2 \leq \Delta$ với $\Delta > 0$ được gọi là bán kính trust region.

m_k thường được chọn là hàm bậc 2

$$m_k(x_k + p) = f(x_k) + p^T \nabla f(x_k) + \frac{1}{2} p^T B_k p$$

với ma trận B_k là ma trận Hessian $\nabla^2 f(x_k)$ hay xấp xỉ của nó.

Line search vs. trust region

Về mặt nào đó có thể coi phương pháp line search và trust region khách nhau về thứ tự chọn hướng và khoảng cách.

Line search vs. trust region

Về mặt nào đó có thể coi phương pháp line search và trust region khách nhau về thứ tự chọn hướng và khoảng cách.

Line search chọn hướng trước rồi theo hướng đó chọn khoảng cách.

Line search vs. trust region

Về mặt nào đó có thể coi phương pháp line search và trust region khách nhau về thứ tự chọn hướng và khoảng cách.

Line search chọn hướng trước rồi theo hướng đó chọn khoảng cách.

Trust region cố định khoảng cách tối đa trước rồi chọn hướng tốt nhất theo ràng buộc đó.

Tài liệu tham khảo

Chương 2, J. Nocedal and S. Wright, Numerical Optimization, Springer.

Chi tiết về Line search: Chương 3, J. Nocedal and S. Wright, Numerical Optimization, Springer.

Chi tiết về Trust region: Chương 4, J. Nocedal and S. Wright, Numerical Optimization, Springer.