

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC KHOA HỌC

HOÀNG NGỌC THẾ

VỀ TỐI ƯU TRÊN ĐA TẠP RIEMANN

LUẬN VĂN THẠC SĨ TOÁN HỌC

Thái Nguyên - 2018

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC KHOA HỌC

HOÀNG NGỌC THẾ

VỀ TỐI ƯU TRÊN ĐA TẠP RIEMANN

Chuyên ngành: Toán ứng dụng

Mã số: 8 46 01 12

LUẬN VĂN THẠC SĨ TOÁN HỌC

Người hướng dẫn khoa học:

TS. NGUYỄN THANH SƠN

Thái Nguyên - 2018

Mục lục

Mở đầu	1
1 Đa tạp và một số khái niệm liên quan	3
1.1 Khái niệm đa tạp	3
1.1.1 Đa tạp khả vi	3
1.1.2 Đa tạp con	6
1.1.3 Vectơ tiếp xúc, không gian tiếp xúc	7
1.1.4 Ánh xạ trên đa tạp	9
1.1.5 Đạo hàm của ánh xạ	10
1.1.6 Một số ánh xạ khả vi đặc biệt	11
1.1.7 Phân thớ tiếp xúc	13
1.1.8 Trường vectơ	13
1.2 Đa tạp Riemann	14
1.2.1 Khái niệm	14
1.2.2 Khoảng cách	15
1.2.3 Gradient	16
1.2.4 Liên thông affine	17
1.2.5 Liên thông Riemann	18
1.2.6 Cung trắc địa, ánh xạ mũ	19
1.2.7 Toán tử Hessian	24
2 Thuật toán Tìm theo đường thẳng trên đa tạp	26

2.1	Thuật toán Tìm theo đường thẳng trong $\mathbb{R}^n$	26
2.1.1	Phương pháp giảm sâu nhất	27
2.1.2	Phương pháp Newton	29
2.2	Tìm theo đường thẳng trên đa tạp Riemann tổng quát	29
2.2.1	Phân tích	29
2.2.2	Thuật toán Tìm theo đường thẳng	32
2.2.3	Sự hội tụ của thuật toán Tìm theo đường thẳng	34
2.2.4	Tốc độ hội tụ	36
2.3	Phương pháp Newton	37
2.3.1	Phương pháp Newton trên đa tạp Riemann với hàm mục tiêu giá trị thực	37
2.3.2	Sự hội tụ địa phương	38
3	Ví dụ về bài toán tối ưu trên mặt cầu	40
3.1	Bài toán K-mean trên mặt cầu	40
3.1.1	Bài toán	40
3.1.2	Thực hành với MATLAB	41
3.2	Bài toán điểm trung chuyển hàng không	47
3.2.1	Giới thiệu	47
3.2.2	Bài toán điểm trung chuyển hàng không	47
3.2.3	Thực hành với MATLAB	49
3.3	Bài toán giá trị riêng dưới góc độ tối ưu	55
3.3.1	Bài toán giá trị riêng dưới góc độ tối ưu	56
3.3.2	Thuật toán thương Rayleigh trên mặt cầu	57
3.3.3	Thực hành với MATLAB	62
	Kết luận	66
	Tài liệu tham khảo	67

Bảng ký hiệu

$\otimes$	tích tenxơ của các không gian vectơ
$C^\infty(\mathcal{M})$	tập tất cả các hàm trơn trên $\mathcal{M}$
$\mathfrak{F}(U)$	tập tất cả hàm trơn, giá trị thực xác định trên U
$\text{grad}f(x)$	Gradient của hàm số f
$\text{Hess}f$	Hessian của hàm số f
S^{m-1}	mặt cầu đơn vị trong $\mathbb{R}^m$
$T_x\mathcal{M}$	không gian tiếp xúc của $\mathcal{M}$ tại x
$\text{tr}(A)$	tổng các phần tử trên đường chéo của ma trận vuông A
$\mathfrak{X}(\mathcal{M})$	tập các trường vectơ trơn trên $\mathcal{M}$.

Mở đầu

Bài toán tối ưu trên đa tạp xuất hiện rất tự nhiên và khá phổ biến. Chẳng hạn, khi xét một bài toán tối ưu có ràng buộc, trong rất nhiều trường hợp, tập chấp nhận được là một đa tạp. Khi đó, ta có một bài toán tối ưu trên đa tạp. Câu hỏi tự nhiên là những phương pháp tối ưu quen thuộc, chẳng hạn Tìm theo đường thẳng, có thể được dùng cho bài toán tương ứng trên đa tạp hay không?

Trong $\mathbb{R}^n$, việc tịnh tiến theo hướng của một vectơ rất rõ ràng: ta chỉ cần cộng vào tọa độ của điểm cần tịnh tiến với tọa độ vectơ tịnh tiến. Nhưng trên một đa tạp, gradient của hàm mục tiêu lại là một vectơ thuộc không gian tiếp xúc của đa tạp chấp nhận được. Khi đó, về nguyên tắc, ta không thể cộng một điểm trên đa tạp với một vectơ trên không gian tiếp xúc. Nếu giả sử có cộng được một cách máy móc thì kết quả đó sẽ không còn là một điểm nằm trên đa tạp nữa. Để góp phần làm sáng tỏ những vấn đề này, chúng tôi đã chọn đề tài "*Về tối ưu trên đa tạp Riemann*" để làm đề tài luận văn thạc sĩ. Nội dung của luận văn được trình bày trong ba chương.

Chương 1. Đa tạp và một số khái niệm liên quan

Trước tiên, chúng tôi trình bày một số khái niệm, tính chất của đa tạp và một số khái niệm liên quan.

Chương 2. Thuật toán Tìm theo đường thẳng trên đa tạp

Trong chương này, chúng tôi trình bày những ý tưởng và xây dựng thuật toán Tìm theo đường thẳng và phương pháp Newton trên đa tạp Riemann. Chúng tôi cũng trình bày một số tính chất về tính hội tụ của thuật toán.

Chương 3. Ví dụ về bài toán tối ưu trên mặt cầu

Trong chương này, chúng tôi áp dụng những thuật toán đã xây dựng cho một số bài toán trên mặt cầu đơn vị S^{n-1} với các ví dụ số cụ thể.

Luận văn kết thúc với phần kết luận và tài liệu tham khảo.

Mặc dù đã rất nghiêm túc và cố gắng thực hiện luận văn này, nhưng luận văn sẽ không tránh khỏi những khiếm khuyết nhất định. Kính mong sự góp ý của các thầy cô để luận văn này được hoàn chỉnh và ý nghĩa hơn.

Luận văn này được thực hiện tại Trường Đại học Khoa học - Đại học Thái Nguyên và hoàn thành dưới sự hướng dẫn của TS. Nguyễn Thanh Sơn. Tác giả xin được bày tỏ lòng biết ơn chân thành và sâu sắc tới người hướng dẫn khoa học của mình, người đã đặt vấn đề nghiên cứu, dành nhiều thời gian hướng dẫn tận tình và đầy trách nhiệm để tác giả hoàn thành luận văn này.

Tác giả đã học tập được rất nhiều kiến thức chuyên ngành bổ ích cho công tác và nghiên cứu của bản thân. Nhân dịp này tác giả xin gửi lời cảm ơn sâu sắc tới các Thầy giáo, Cô giáo đã tham gia giảng dạy lớp Cao học Toán K10A; Nhà trường và các phòng chức năng của Trường, Khoa Toán - Tin, trường Đại học Khoa học, Đại học Thái Nguyên đã quan tâm và giúp đỡ tác giả trong suốt thời gian học tập tại trường.

Cuối cùng tác giả xin cảm ơn gia đình, bạn bè và đồng nghiệp đã động viên, ủng hộ và tạo mọi điều kiện cho tác giả trong suốt thời gian nghiên cứu và học tập.

Thái Nguyên, tháng 9 năm 2018

Tác giả luận văn

Hoàng Ngọc Thế

Chương 1

Đa tập và một số khái niệm liên quan

Trong chương này, chúng tôi trình bày một cách chi tiết những khái niệm, tính chất quan trọng liên quan đến đa tập. Những nội dung này chủ yếu dựa vào tài liệu [1]. Người đọc cũng có thể tham khảo tài liệu [2].

1.1 Khái niệm đa tập

1.1.1 Đa tập khả vi

Định nghĩa 1.1.1. Giả sử $(\mathcal{M}, \tau)$ là không gian tô pô Hausdorff với một cơ sở đếm được. Khi đó $\mathcal{M}$ được gọi là một *đa tập tô pô m -chiều* nếu nó đồng phôi địa phương với không gian $\mathbb{R}^m$, nghĩa là với mỗi điểm $x \in \mathcal{M}$, tồn tại một lân cận U của x , có một tập con mở $V \subset \mathbb{R}^m$ và một phép đồng phôi $\varphi : U \rightarrow V$.

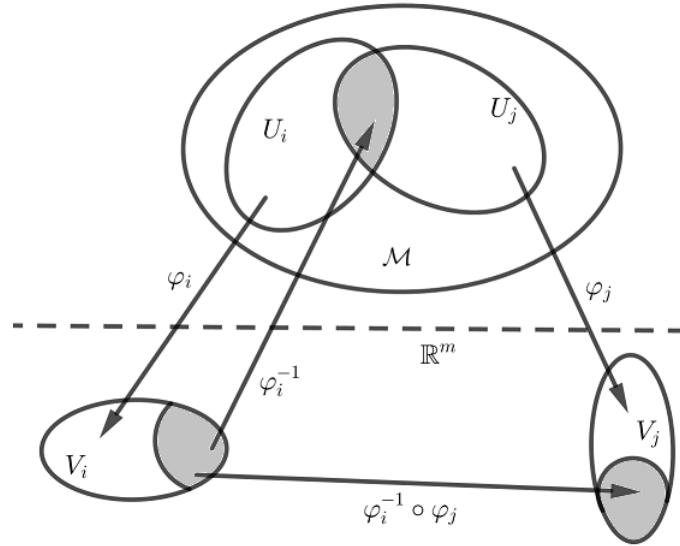
Cặp (U, φ) được gọi là một *bản đồ địa phương* hay gọi tắt là *bản đồ* trong $\mathcal{M}$.

Ta viết $\mathcal{M}^m$ để thể hiện đa tập $\mathcal{M}$ có m chiều.

Với U là một tập mở bất kỳ của $\mathbb{R}^n$, ta nhắc lại các kí hiệu $C^k(U, \mathbb{R}^m)$, $C^\infty(U, \mathbb{R}^m)$, và $C^\omega(U, \mathbb{R}^m)$ lần lượt là tập tất cả các ánh xạ khả vi liên tục tới cấp k , tập tất cả các ánh xạ trơn và tập tất cả các ánh xạ giải tích từ U vào $\mathbb{R}^m$.

Định nghĩa 1.1.2. Xét đa tập tô pô $\mathcal{M}^m$. Họ $\mathfrak{A} = \{(U_i, \varphi_i) : i \in I\}$ các bản đồ trên $\mathcal{M}$ được gọi là một *atlas lớp C^k* ($k \geq 1$) hay *C^k -atlas* nếu hai điều kiện sau được thỏa mãn:

- (i) họ $\{U_i\}$ là một phủ mở của $\mathcal{M}$;
- (ii) với hai bản đồ (U_i, φ_i) và (U_j, φ_j) mà $U_i \cap U_j \neq \emptyset$ thì ánh xạ chuyển tiếp $\varphi_j \circ \varphi_i^{-1}$ xác định trên $\varphi_i(U_i \cap U_j)$ là ánh xạ khả vi lớp C^k từ $\varphi_i(U_i \cap U_j)$ lên $\varphi_j(U_i \cap U_j)$.



Một bản đồ (U, φ) được gọi là *tương thích* với C^k – atlas $\mathfrak{A}$ nếu hợp $\mathfrak{A} \cup \{(U, \varphi)\}$ là một C^k – atlas.

Atlas $\hat{\mathfrak{A}}$ được gọi là *atlas cực đại* nếu nó chứa tất cả các bản đồ tương thích với nó. Khi đó $\hat{\mathfrak{A}}$ cũng được gọi là một C^k – *cấu trúc* trên $\mathcal{M}^m$.

Cặp $(\mathcal{M}, \hat{\mathfrak{A}})$ được gọi là một C^k – *đa tạp* hay *đa tạp khả vi lớp C^k* .

Nhận xét 1.1.3. Một C^k – atlas $\mathfrak{A}$ trên một đa tạp tô pô $\mathcal{M}$ xác định duy nhất một C^k – cấu trúc trên $\mathcal{M}$.

Ví dụ 1.1.4.

1. Xét $\mathcal{M} = \mathbb{R}^m$ với tô pô Euclide. Ta có một C^ω – cấu trúc tầm thường

$$\mathfrak{A} = \{(\mathbb{R}^m, \varphi) : \varphi : x \mapsto x\}.$$

2. Xét không gian vectơ $\mathbb{R}^{n \times p}$ các ma trận thực cỡ $n \times p$. Ta xây dựng tích vô hướng

$$\langle X_1, X_2 \rangle := \text{tr} (X_1^\top X_2),$$

ở đây $\text{tr}(X)$ là tổng các phần tử trên đường chéo chính của X . Chuẩn tương ứng được gọi là *chuẩn Frobenius*

$$\|X\|_F = \sqrt{\text{tr} (X^\top X)},$$

tức là $\|X\|_F^2$ là tổng bình phương các phần tử của X .

Khi đó $\mathbb{R}^{n \times p}$ là một không gian tô pô Hausdorff với cơ sở đếm được.

Xét ánh xạ

$$\begin{aligned} \text{vec} : \mathbb{R}^{n \times p} &\rightarrow \mathbb{R}^{np} \\ X &\mapsto \text{vec}(X), \end{aligned}$$

ở đây $\text{vec}(X)$ là vectơ được tạo thành bằng cách xếp chồng theo thứ tự các cột của X . Ánh xạ này cho ta một C^k -cấu trúc trên $\mathbb{R}^{n \times p}$.

3. Ký hiệu S^m là mặt cầu đơn vị trong $\mathbb{R}^{m+1}$ được trang bị tô pô Euclide. Ký hiệu $N = (0, 0, \dots, 0, 1)$, $S = (0, 0, \dots, 0, -1)$ lần lượt là cực bắc và cực nam của S^m . Đặt

$$U_N = S^m \setminus \{N\}, U_S = S^m \setminus \{S\}$$

và định nghĩa

$$\begin{aligned} \varphi_N : U_N &\rightarrow \mathbb{R}^m \\ (x_1, \dots, x_{m+1}) &\mapsto \left(\frac{x_1}{1 - x_{m+1}}, \dots, \frac{x_m}{1 - x_{m+1}} \right), \\ \varphi_S : U_S &\rightarrow \mathbb{R}^m \\ (x_1, \dots, x_{m+1}) &\mapsto \left(\frac{x_1}{1 + x_{m+1}}, \dots, \frac{x_m}{1 + x_{m+1}} \right). \end{aligned}$$

Do vậy, hai ánh xạ chuyển tiếp $\varphi_S^{-1} \varphi_N$ và $\varphi_N \varphi_S^{-1}$ được cho bởi công thức

$$\begin{aligned} \mathbb{R}^m \setminus \{0\} &\rightarrow \mathbb{R}^m \setminus \{0\} \\ x &\mapsto \frac{x}{\|x\|^2}. \end{aligned}$$

Khi đó $\mathfrak{A} = \{(U_N, \varphi_N), (U_S, \varphi_S)\}$ là một C^ω -Atlas trên S^m . Đa tập lớp $C^\infty(S^m, \hat{\mathfrak{A}})$ được gọi là *mặt cầu tiêu chuẩn* m chiều.

Phát biểu sau đây xây dựng tích Descartes của hai đa tập.

Mệnh đề 1.1.5. Cho $(\mathcal{M}_1, \hat{\mathfrak{A}}_1)$ và $(\mathcal{M}_2, \hat{\mathfrak{A}}_2)$ là hai đa tập khả vi lớp C^k . Cho $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$ là tích Descartes của hai không gian tô pô. Khi đó, tồn tại một atlas $\mathfrak{A}$ trên $\mathcal{M}$ sao cho $(\mathcal{M}, \hat{\mathfrak{A}})$ là một đa tập khả vi lớp C^k có số chiều

$$\dim \mathcal{M} = \dim \mathcal{M}_1 + \dim \mathcal{M}_2.$$

Đa tập $(\mathcal{M}, \hat{\mathfrak{A}})$ được gọi là *đa tập tích Descartes* của hai đa tập $(\mathcal{M}_1, \hat{\mathfrak{A}}_1)$ và $(\mathcal{M}_2, \hat{\mathfrak{A}}_2)$.

1.1.2 Đa tập con

Định nghĩa 1.1.6. Cho $m \leq n$ là các số nguyên dương và $(\mathcal{N}^n, \hat{\mathfrak{A}}_{\mathcal{N}})$ là một đa tập khả vi lớp C^k . Một tập con $\mathcal{M}$ của $\mathcal{N}$ được gọi là một *đa tập con* của $\mathcal{N}$ nếu với mỗi điểm $x \in \mathcal{M}$, tồn tại một bản đồ $(U_x, \varphi_x) \in \hat{\mathfrak{A}}_{\mathcal{N}}$ sao cho $x \in U_x$ và $\varphi_x : U_x \subset \mathcal{N} \rightarrow \mathbb{R}^m \times \mathbb{R}^{n-m}$ thỏa mãn

$$\varphi_x(U_x \cap \mathcal{M}) = \varphi_x(U_x) \cap (\mathbb{R}^m \times \{0\}),$$

số $n - m$ được gọi là *đối chiều* của $\mathcal{M}$ trong $\mathcal{N}$.

Phát biểu sau đây cung cấp chi tiết hơn về cấu trúc khả vi trên đa tập con.

Mệnh đề 1.1.7. Cho $m \leq n$ là các số nguyên dương và $(\mathcal{N}^n, \hat{\mathfrak{A}}_{\mathcal{N}})$ là một đa tập khả vi lớp C^k . Cho $\mathcal{M}$ là một đa tập con của $\mathcal{N}$ và được trang bị tô pô con. Ký hiệu

$$\pi : \mathbb{R}^m \times \mathbb{R}^{n-m} \rightarrow \mathbb{R}^m$$

là phép chiếu lên thành phần thứ nhất. Khi đó

$$\mathfrak{A}_{\mathcal{M}} := \{(U_x \cap \mathcal{M}, (\pi \circ \varphi_x)|_{U_x \cap \mathcal{M}}) : x \in \mathcal{M}\}$$

là một atlas lớp C^k trên $\mathcal{M}$. Vì thế $(\mathcal{M}, \hat{\mathfrak{A}}_{\mathcal{M}})$ là một đa tập khả vi m chiều lớp C^k .

Cấu trúc khả vi $\mathfrak{A}_{\mathcal{M}}$ trên $\mathcal{M}$ xác định trong mệnh đề trên được gọi là *cấu trúc cảm sinh* từ $\mathfrak{A}_{\mathcal{N}}$.

Ta nhắc lại rằng một ma trận A cỡ $m \times n$ được gọi là *hạng đủ* nếu $\text{rank} A = \min\{m, n\}$, một ánh xạ tuyến tính giữa các không gian hữu hạn chiều đều có thể coi tương đương như một ma trận. Định lý sau đây cho ta một nguồn dồi dào những ví dụ về đa tạp con.

Định lý 1.1.8 (Ánh xạ ẩn). Cho $m \leq n$ là các số nguyên dương và $f : U \rightarrow \mathbb{R}^m$ là một ánh xạ lớp C^k từ một tập mở $U \subset \mathbb{R}^n$. Nếu $x \in U$, $f(x) = y$ và Jacobian của nó

$$df|_x : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

là hạng đủ thì $f^{-1}(\{y\})$ là một đa tạp con khả vi lớp C^k của $\mathbb{R}^n$ có số chiều $n - m$.

Ví dụ 1.1.9. Cho

$$\begin{aligned} f : \mathbb{R}^{m+1} &\rightarrow \mathbb{R} \\ x &\mapsto \sum_{i=1}^{m+1} x_i^2. \end{aligned}$$

Dễ thấy f thuộc lớp C^ω . Khi đó, đạo hàm của f được tính bằng $df_x = 2x$. Có thể thấy ngay $1 \in \mathbb{R}$ là giá trị thỏa mãn điều kiện của định lý ánh xạ ẩn và vì thế

$$S^m := \{x \in \mathbb{R}^{n+1} : \|x\|^2 = 1\} = f^{-1}(\{1\})$$

là một đa tạp lớp C^ω và có số chiều m . Đây chính là mặt cầu tiêu chuẩn đã được nêu ở Ví dụ 1.1.4.

1.1.3 Vectơ tiếp xúc, không gian tiếp xúc

Trước tiên, ta định nghĩa khái niệm này trong không gian $\mathbb{R}^n$ quen thuộc.

Định nghĩa 1.1.10. Cho x là một điểm trong $\mathbb{R}^m$ và ký hiệu $T_x \mathbb{R}^m$ là tập các toán tử vi phân tuyến tính tại x triệt tiêu hằng số. Tức là $T_x \mathbb{R}^m$ gồm các ánh xạ $\alpha : \varepsilon(x) \rightarrow \mathbb{R}$ thỏa mãn

$$(i) \alpha(\lambda f + \mu g) = \lambda \alpha(f) + \mu \alpha(g),$$

$$(ii) \alpha(fg) = \alpha(f) \cdot g(x) + f(x)\alpha(g),$$

với mọi $\alpha, \mu \in \mathbb{R}$ và $f, g \in \varepsilon(x)$.

Để dàng nhận thấy tập $T_x \mathbb{R}^m$ có cấu trúc của một không gian vectơ thực với hai phép toán

$$(\alpha + \mu)(f) := \alpha(f) + \mu(f),$$

$$(\lambda \alpha)(f) := \lambda \alpha(f).$$

Định lí 1.1.11. Cho $x \in \mathbb{R}^m$. Khi đó, ánh xạ

$$\begin{aligned} \phi : \mathbb{R}^m &\rightarrow T_x \mathbb{R}^m \\ v &\mapsto \partial_v \end{aligned}$$

là một đẳng cấu giữa các không gian vectơ.

Bây giờ ta xét các khái niệm đó đối với đa tạp.

Định nghĩa 1.1.12. Cho $\mathcal{M}$ là một đa tạp khả vi, $x \in \mathcal{M}$ và $\varepsilon(x)$ là tập các hàm thực định nghĩa trên một lân cận mở của x . Một *vectơ tiếp xúc* ξ_x tại x là một ánh xạ $\xi_x : \varepsilon(x) \rightarrow \mathbb{R}$ thỏa mãn

$$(i) \xi_x(\lambda f + \mu g) = \lambda \xi_x(f) + \mu \xi_x(g),$$

$$(ii) \xi_x(fg) = \xi_x(f)g(x) + f(x)\xi_x(g),$$

với mọi $\lambda, \mu \in \mathbb{R}$ và $f, g \in \varepsilon(x)$.

Tập các vectơ tiếp xúc của $\mathcal{M}$ tại x được gọi là *không gian tiếp xúc* tại x và ký hiệu $T_x \mathcal{M}$.

Phép cộng và phép nhân với vô hướng trong $T_x \mathcal{M}$ được định nghĩa như sau

$$(\xi_x + \zeta_x)(f) = \xi_x(f) + \zeta_x(f),$$

$$(\lambda \xi_x)(f) = \lambda \xi_x(f),$$

với mọi $\xi_x, \zeta_x \in T_x \mathcal{M}$, $f \in \varepsilon(x)$ và $\lambda \in \mathbb{R}$.

Ví dụ 1.1.13. Cho $\gamma : I \rightarrow S^m$ là một cung trên mặt cầu đơn vị trong $\mathbb{R}^{m+1}$ sao cho $\dot{\gamma}(0) = \xi$. Do $\gamma(t)$ nằm trên S^m nên

$$\gamma(t)^\top \gamma(t) = 1, \forall t \in I.$$

Đạo hàm hai vế cho ta

$$\dot{\gamma}(t)^\top \gamma(t) = \gamma(t)^\top \dot{\gamma}(t) = 0,$$

hay

$$\dot{\gamma}(t)^\top \gamma(t) = 0.$$

Từ đó, ta suy ra rằng với $x \in S^m$, ξ là vectơ tiếp xúc x thì $\xi^\top x = 0$ hay $\forall \xi \in T_x S^m$, $\xi \perp x$. Do vậy, ta viết

$$T_x S^m = \{ \xi \in \mathbb{R}^{m+1} : \xi^\top x = 0 \}.$$

1.1.4 Ánh xạ trên đa tạp

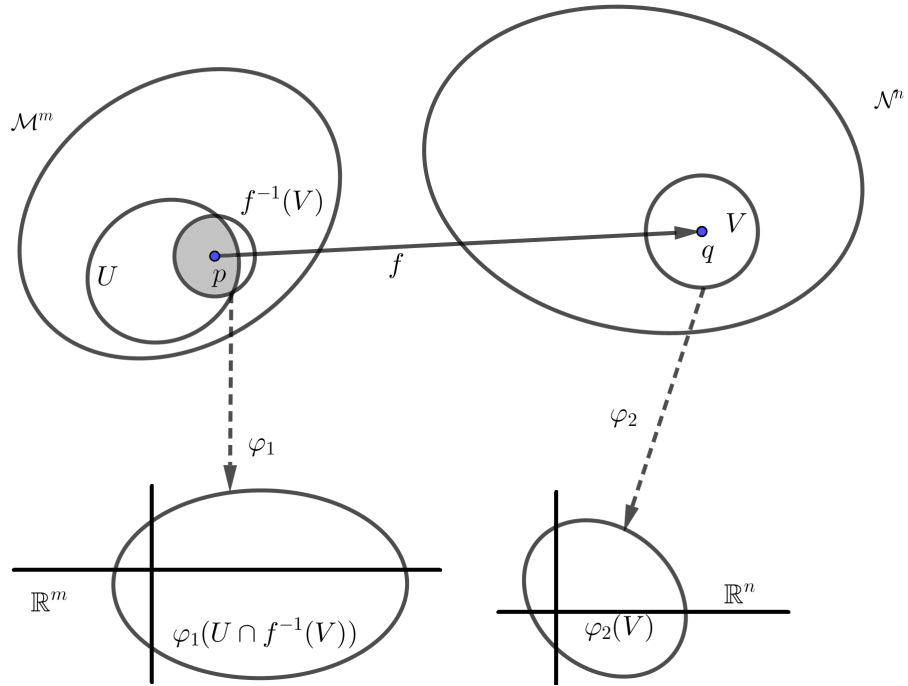
Định nghĩa 1.1.14. Cho $(\mathcal{M}^m, \mathfrak{A}_{\mathcal{M}})$ và $(\mathcal{N}^n, \mathfrak{A}_{\mathcal{N}})$ là các đa tạp khả vi lớp C^k . Ánh xạ $f : \mathcal{M} \rightarrow \mathcal{N}$ được gọi là *khả vi lớp C^k* nếu với mỗi bản đồ $(U, \varphi_1) \in \mathfrak{A}_{\mathcal{M}}$ và $(V, \varphi_2) \in \mathfrak{A}_{\mathcal{N}}$, ánh xạ

$$\varphi_2 \circ f \circ \varphi_1^{-1} : \varphi_1(U \cap f^{-1}(V)) \subset \mathbb{R}^m \rightarrow \mathbb{R}^n$$

thuộc lớp C^k .

Ánh xạ khả vi $\gamma : I \rightarrow \mathcal{M}$ xác định trên khoảng mở $I \subset \mathbb{R}$ được gọi là một *cung khả vi* trong $\mathcal{M}$.

Một ánh xạ khả vi $f : \mathcal{M} \rightarrow \mathbb{R}$ được gọi là một *hàm số khả vi* trên $\mathcal{M}$. Tập tất cả các hàm trơn trên $\mathcal{M}$ được ký hiệu là $C^\infty(\mathcal{M})$.



1.1.5 Đạo hàm của ánh xạ

Định nghĩa 1.1.15. Cho $\phi : \mathcal{M} \rightarrow \mathcal{N}$ là một ánh xạ khả vi giữa các đa tạp khả vi. Đạo hàm $D\phi(x)$ của ϕ tại x trong $\mathcal{M}$ là ánh xạ $D\phi(x) : T_x\mathcal{M} \rightarrow T_{\phi(x)}\mathcal{N}$ sao cho mỗi $\xi_x \in T_x\mathcal{M}$ và $f \in \varepsilon(\phi(x))$, ta có

$$D\phi(x)[\xi_x] = \xi_x(f \circ \phi).$$

Giả sử $\gamma : I \rightarrow \mathcal{M}$ là một cung trên $\mathcal{M}$ sao cho $\gamma(0) = x$ và $\dot{\gamma}(0) = \xi_x$. Ký hiệu $c : I \rightarrow \mathcal{N}$ là ảnh của cung γ qua ϕ , tức là $c = \phi \circ \gamma$ với $\dot{c}(0) = \phi(x)$ và đặt $\zeta_{\phi(x)} = \dot{c}(0)$. Khi đó, với $f \in \varepsilon(\phi(x))$, ta có

$$\begin{aligned} (D\phi(x)[\xi_x])(f) &= \xi_x(f \circ \phi) \\ &= \frac{d}{dt}(f \circ \phi \circ \gamma(t))|_{t=0} \\ &= \frac{d}{dt}(f \circ c(t))|_{t=0} \\ &= \zeta_{\phi(x)}(f). \end{aligned}$$

Đặc biệt, nếu $\mathcal{M}$ và $\mathcal{N}$ là các đa tạp tuyến tính thì $DF(x)$ chính là đạo hàm theo

nghĩa cổ điển

$$DF(x)[\xi_x] = \lim_{t \rightarrow 0} \frac{F(x + t\xi_x) - F(x)}{t}. \quad (1.1)$$

Mệnh đề 1.1.16. Cho $\phi : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ và $\psi : \mathcal{M}_2 \rightarrow \mathcal{M}_3$ là các ánh xạ khả vi giữa các đa tạp khả vi. Khi đó, với mỗi điểm $x \in \mathcal{M}_1$, ta có

- (i) Ánh xạ $D\phi(x) : T_x\mathcal{M}_1 \rightarrow T_{\phi(x)}\mathcal{M}_2$ là tuyến tính;
- (ii) $D(\text{id}_{\mathcal{M}_1})(x) = \text{id}_{T_x\mathcal{M}_1}$;
- (iii) $D(\psi \circ \phi)(x) = D\psi(\phi(x)) \circ D\phi(x)$.

Một hệ quả trực tiếp của mệnh đề trên là.

Hệ quả 1.1.17. Cho $\phi : \mathcal{M} \rightarrow \mathcal{N}$ là một vi phôi với ánh xạ ngược $\phi^{-1} : \mathcal{N} \rightarrow \mathcal{M}$. Khi đó, đạo hàm $D\phi(x) : T_x\mathcal{M} \rightarrow T_{\phi(x)}\mathcal{N}$ tại x là một song ánh và

$$(D\phi(x))^{-1} = D\phi^{-1}(\phi(x)). \quad (1.2)$$

Định lý sau đây cung cấp cho ta thông tin về số chiều của không gian tiếp xúc.

Định lý 1.1.18. Cho $\mathcal{M}^m$ là một đa tạp khả vi m chiều và $x \in \mathcal{M}$. Khi đó, không gian tiếp xúc $T_x\mathcal{M}$ của $\mathcal{M}$ tại x là một không gian vectơ m chiều.

1.1.6 Một số ánh xạ khả vi đặc biệt

Định nghĩa 1.1.19. Cho $\phi : \mathcal{M}^m \rightarrow \mathcal{N}^n$ là một ánh xạ khả vi giữa các đa tạp khả vi,

- (i) với $m \leq n$, ϕ được gọi là một *dìm* nếu với mỗi $x \in \mathcal{M}$, đạo hàm của ϕ tại x , $D\phi(x) : T_x\mathcal{M} \rightarrow T_{\phi(x)}\mathcal{N}$ là một đơn ánh;
- (ii) ϕ được gọi là một *nhúng* nếu nó là một dìm và đồng thời là một đồng phôi lên $\phi(\mathcal{M})$.
- (iii) với $m \geq n$, ϕ được gọi là 1 *ngập* nếu với mỗi $x \in \mathcal{M}$, $D\phi_x$ là một toàn ánh.

Ví dụ 1.1.20.

1. Phép nhúng thông thường từ $\mathbb{R}^{m+1}$ vào $\mathbb{R}^{n+1}$ là một đim, đồng thời là một nhúng. Ngoài ra, nó cũng là một nhúng từ S^m vào S^n .
2. Phép chiếu chính tắc từ $\mathbb{R}^{m+1}$ lên $\mathbb{R}^{n+1}$, $m \geq n$ là một ngập.

Định nghĩa 1.1.21. Cho $\mathcal{M}$ là một đa tạp khả vi m chiều và U là một tập mở trong $\mathbb{R}^m$. Một đim $\phi : U \rightarrow \mathcal{M}$ được gọi là một *tham số hóa địa phương* của $\mathcal{M}$. Nếu ϕ đồng thời là một toàn ánh thì nó được gọi là một *tham số hóa toàn cục*.

Kết quả sau đây mở rộng Định lý Ánh xạ ngược.

Định lý 1.1.22. Cho $\phi : \mathcal{M} \rightarrow \mathcal{N}$ là một ánh xạ khả vi giữa các đa tạp cùng chiều. Nếu $x \in \mathcal{M}$ thỏa mãn $D\phi(x) : T_x\mathcal{M} \rightarrow T_{\phi(x)}\mathcal{N}$ là song ánh thì tồn tại một lân cận mở U_x của x và U_y của $y = \phi(x)$ sao cho $\psi = \phi|_{U_x} : U_x \rightarrow U_y$ là song ánh và ánh xạ ngược $\psi^{-1} : U_y \rightarrow U_x$ cũng khả vi.

Tiếp theo, ta sẽ mở rộng Định lý ánh xạ ẩn. Trước tiên ta sẽ bổ sung một số khái niệm.

Định nghĩa 1.1.23. Cho $\phi : \mathcal{M}^m \rightarrow \mathcal{N}^n$ là một ánh xạ khả vi giữa các đa tạp. Một điểm $x \in \mathcal{M}$ được gọi là *chính quy* nếu đạo hàm $D\phi(x) : T_x\mathcal{M} \rightarrow T_{\phi(x)}\mathcal{N}$ là hạng đủ. Ngược lại, ta gọi x là *điểm dị thường*. Một điểm $y \in \phi(\mathcal{M})$ được gọi là một *giá trị chính quy* nếu $\forall x \in \phi^{-1}\{y\}$, x là điểm chính quy.

Định lý 1.1.24. Cho $\phi : \mathcal{M}^m \rightarrow \mathcal{N}^n$ là một ánh xạ khả vi giữa các đa tạp với $m > n$. Nếu $y \in \phi(\mathcal{M})$ là một giá trị chính quy thì nghịch ảnh của nó, $\phi^{-1}(\{y\})$ là một đa tạp con của $\mathcal{M}^m$ có số chiều $m - n$. Không gian tiếp xúc $T_x\phi^{-1}(\{y\})$ của $\phi^{-1}(\{y\})$ tại x là nhân của đạo hàm ánh xạ $D\phi(x)$, tức là

$$T_x\phi^{-1}(\{y\}) = \{\xi \in T_x\mathcal{M} : D\phi(x)[\xi] = 0\}.$$

1.1.7 Phân thớ tiếp xúc

Cho đa tạp $\mathcal{M}$, kí hiệu $T\mathcal{M}$ là tập tất cả không gian tiếp xúc với $\mathcal{M}$

$$T\mathcal{M} := \bigcup_{x \in \mathcal{M}} T_x\mathcal{M}.$$

Vì với mỗi $\xi \in T\mathcal{M}$, có một và chỉ một không gian tiếp xúc $T_x\mathcal{M}$, nên $\mathcal{M}$ là một thương của $T\mathcal{M}$ với phép chiếu tự nhiên

$$\pi : T\mathcal{M} \rightarrow \mathcal{M}$$

$$\xi \in T_x\mathcal{M} \mapsto x.$$

Xét bản đồ (U, φ) của $\mathcal{M}$, ánh xạ

$$\xi \in T_x\mathcal{M} \mapsto (\varphi_1(x), \dots, \varphi_m(x), \xi_{\varphi_1}, \dots, \xi_{\varphi_m})^\top,$$

người ta chứng minh được ánh xạ trên cho ta một cấu trúc đa tạp trên $T\mathcal{M}$. Ta gọi $T\mathcal{M}$ là một *phân thớ tiếp xúc* của $\mathcal{M}$.

1.1.8 Trường vectơ

Ta nhắc lại khái niệm trường vectơ dọc một cung tham số trong $\mathbb{R}^n$. Trường vectơ dọc cung tham số $\gamma : I \rightarrow \mathbb{R}^n$ là ánh xạ $\xi : I \rightarrow T\mathbb{R}^n$ sao cho $\xi(t) \in T_{\gamma(t)}\mathbb{R}^n$.

Ta khái quát hóa khái niệm này cho đa tạp.

Định nghĩa 1.1.25. Cho đa tạp $\mathcal{M}$ và một cung tham số $\gamma : I \rightarrow \mathcal{M}$. Ánh xạ trơn đặt tương ứng mỗi $t \in I$, một vectơ tiếp xúc với $\mathcal{M}$ tại $\gamma(t)$ được gọi là một *trường vectơ dọc cung* γ .

Ví dụ 1.1.26. Dễ thấy $t \mapsto \gamma \xi_{\gamma(t)}$ và $t \mapsto \dot{\gamma}(t)$ là những trường vectơ dọc cung γ .

Định nghĩa 1.1.27. Cho đa tạp $\mathcal{M}$. Ánh xạ trơn

$$\xi : \mathcal{M} \rightarrow T\mathcal{M}$$

$$x \mapsto \xi_x \in T_x\mathcal{M}$$

được gọi là một *trường vectơ* trên đa tạp $\mathcal{M}$.

Cho một trường vectơ ξ trên đa tạp $\mathcal{M}$ và một hàm (giá trị thực) trơn $f \in \mathfrak{F}(\mathcal{M})$, phép cộng hai trường vectơ và phép nhân một trường vectơ với một hàm $f \in \mathfrak{F}(\mathcal{M})$ được định nghĩa như sau

$$\begin{aligned}(f\xi)_x &:= f(x)\xi_x, \\ (\xi + \zeta)_x &:= \xi_x + \zeta_x, \quad \forall x \in \mathcal{M}.\end{aligned}$$

Tập tất cả trường vectơ trơn trên $\mathcal{M}$ được trang bị hai phép toán trên được kí hiệu là $\mathfrak{X}(\mathcal{M})$.

Với mỗi $\xi \in \mathfrak{X}(\mathcal{M})$ và hàm khả vi $f \in \mathfrak{F}_x(\mathcal{M})$, ta xác định hàm $\xi(f) \in \varepsilon(x)$ như sau: với $x \in \mathcal{M}$,

$$(\xi(f))(x) = \xi_x(f) = \left. \frac{d(f(\gamma(t)))}{dt} \right|_{t=0},$$

trong đó $\gamma(0) = x$.

1.2 Đa tạp Riemann

1.2.1 Khái niệm

Cho $\mathcal{M}$ là một đa tạp trơn, tập các hàm trơn $C^\infty(\mathcal{M})$ trên $\mathcal{M}$ lập thành một vành giao hoán. Tập các hàm trơn $C^\infty(T\mathcal{M})$ trên $\mathcal{M}$ lập thành một môđun trên $C^\infty(\mathcal{M})$. Đặt $C_0^\infty(T\mathcal{M}) = C^\infty(\mathcal{M})$ và với số nguyên dương r , đặt

$$C_r^\infty(T\mathcal{M}) = C^\infty(T\mathcal{M}) \otimes \dots \otimes C^\infty(T\mathcal{M}),$$

trong đó $\otimes$ là ký hiệu tích tenxơ của các không gian vectơ.

Định nghĩa 1.2.1. Cho $\mathcal{M}$ là một đa tạp khả vi. Một *mêtric Riemann* g trên $\mathcal{M}$ là một ánh xạ song tuyến tính : $C_2^\infty(T\mathcal{M}) \rightarrow C_0^\infty(T\mathcal{M})$ sao cho với mỗi $x \in \mathcal{M}$, hạn chế g_x của g trên $T_x\mathcal{M} \otimes T_x\mathcal{M}$ với

$$g_x(\xi_x, \zeta_x) = g(\xi, \zeta)_x = \langle \xi_x, \zeta_x \rangle_x.$$

là một tích vô hướng trên không gian tiếp xúc $T_x\mathcal{M}$. Cặp $(\mathcal{M}, g)$ được gọi là *một đa tạp Riemann*.

Định nghĩa 1.2.2. Cho $\mathcal{M}$ là một đa tạp con của đa tạp Riemann $(\overline{\mathcal{M}}, \overline{g})$. Khi đó mọi không gian tiếp xúc $T_x\mathcal{M}$ đều có thể coi như là một không gian con của $T_x\overline{\mathcal{M}}$. Metric Riemann trong $\mathcal{M}$ có thể được định nghĩa như sau

$$g_x(\xi, \zeta) := \overline{g}_x(\xi, \zeta), \quad \xi, \zeta \in T_x\mathcal{M},$$

trong đó ξ và ζ ở vế phải được xem như là các phần tử của $T_x\overline{\mathcal{M}}$. Được trang bị metric như trên, $\mathcal{M}$ trở thành một đa tạp Riemann, gọi là *đa tạp Riemann con* của $\overline{\mathcal{M}}$.

Phần bù trực giao của $T_x\mathcal{M}$ trong $T_x\overline{\mathcal{M}}$ được gọi là *không gian pháp tuyến* của $\mathcal{M}$ tại x :

$$(T_x\mathcal{M})^\perp = \{\xi \in T_x\overline{\mathcal{M}} : g_x(\xi, \zeta) = 0, \forall \zeta \in T_x\mathcal{M}\}.$$

Mọi phần tử $\xi \in T_x\overline{\mathcal{M}}$ đều có thể biểu diễn duy nhất dưới dạng tổng của một phần tử của $T_x\mathcal{M}$ và một phần tử của $(T_x\mathcal{M})^\perp$:

$$\xi = P_x\xi + P_x^\perp\xi,$$

trong đó P_x là phép chiếu trực giao lên $T_x\mathcal{M}$ và $P_x^\perp$ là phép chiếu trực giao lên $(T_x\mathcal{M})^\perp$.

Metric Riemann cho phép ta lượng hóa, đo đạc trên các không gian tiếp xúc. Hơn thế nữa, nó giúp xây dựng cấu trúc không gian metric từ cấu trúc hình học trên đa tạp Riemann như khái niệm khoảng cách trong mục sau đây.

1.2.2 Khoảng cách

Định nghĩa 1.2.3. Cho $(\mathcal{M}, g)$ là một đa tạp Riemann và $\gamma : I \rightarrow \mathcal{M}$ là một cung lớp C^1 trong $\mathcal{M}$. Khi đó, độ dài của γ được định nghĩa là

$$L(\gamma) = \int_I \sqrt{g(\dot{\gamma}(t), \dot{\gamma}(t))} dt. \quad (1.3)$$

Mệnh đề 1.2.4. Cho $(\mathcal{M}, g)$ là một đa tạp Riemann liên thông đường. Với mỗi điểm $x, y \in \mathcal{M}$, ký hiệu $C_{\mathcal{M}}$ là tập tất cả các cung $\gamma : [0, 1] \rightarrow \mathcal{M}$ sao cho $\gamma(0) = x, \gamma(1) = y$

và định nghĩa hàm

$$d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}^+$$

$$(x, y) \mapsto \inf\{L(\gamma) : \gamma \in C_{\mathcal{M}}\}.$$

Khi đó, $(\mathcal{M}, d)$ lập thành một không gian metric. Hơn nữa, tô pô sinh bởi d trùng với tô pô ban đầu của $\mathcal{M}$.

Ta có định lý quan trọng sau đây về sự tồn tại của metric Riemann.

Định lý 1.2.5. Cho $(\mathcal{M}^m, \mathfrak{A})$ là một đa tạp khả vi. Khi đó, tồn tại một metric Riemann trên $\mathcal{M}$.

1.2.3 Gradient

Định nghĩa 1.2.6. Cho một hàm trơn f trên đa tạp Riemann $\mathcal{M}$. Gradient của f tại x , kí hiệu $\text{grad}f(x)$, là phần tử duy nhất của $T_x\mathcal{M}$ thỏa mãn

$$\langle \text{grad}f(x), \xi \rangle = Df(x)[\xi], \quad \forall \xi \in T_x\mathcal{M}. \quad (1.4)$$

Trong thực hành tính toán, rất nhiều trường hợp khi $\mathcal{M}$ là một đa tạp Riemann con của $\mathbb{R}^n$, người ta hay sử dụng phép chiếu trực giao để xác định gradient.

Mệnh đề 1.2.7. Giả sử $\bar{f}$ là một hàm mục tiêu trên đa tạp Riemann $\overline{\mathcal{M}}$ và f là hạn chế của $\bar{f}$ trên đa tạp Riemann con $\mathcal{M}$. Khi đó $\text{grad}f$ là ảnh của $\text{grad}\bar{f}$ qua phép chiếu trực giao lên không gian $T_x\mathcal{M}$.

$$\text{grad}f(x) = P_x \text{grad}\bar{f}(x). \quad (1.5)$$

Chứng minh. Thật vậy, $P_x \text{grad}\bar{f}(x) \in T_x\mathcal{M}$ và với mọi $\zeta \in T_x\mathcal{M}$, ta có

$$\begin{aligned} \langle P_x \text{grad}\bar{f}(x), \zeta \rangle &= \langle \text{grad}\bar{f}(x) - P_x^\perp \text{grad}\bar{f}(x), \zeta \rangle \\ &= \langle \text{grad}\bar{f}(x), \zeta \rangle \\ &= D\bar{f}(x)[\zeta] = Df(x)[\zeta]. \end{aligned}$$

Vậy (1.4) được thỏa mãn. □

1.2.4 Liên thông affine

Định nghĩa 1.2.8. Kí hiệu $\mathfrak{X}(\mathcal{M})$ là tập các trường vectơ trơn trên $\mathcal{M}$. Ánh xạ

$$\begin{aligned}\nabla : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) &\rightarrow \mathfrak{X}(\mathcal{M}) \\ (\eta, \xi) &\mapsto \nabla_\eta \xi\end{aligned}$$

được gọi là một *liên thông affine* trên đa tạp $\mathcal{M}$ nếu các điều kiện sau được thỏa mãn với mọi $\eta, \chi, \xi, \zeta \in \mathfrak{X}(\mathcal{M})$; $f, g \in \mathfrak{F}(\mathcal{M})$; $a, b \in \mathbb{R}$

- (i) $\nabla_{f\eta+g\chi}\xi = f\nabla_\eta\xi + g\nabla_\chi\xi$,
- (ii) $\nabla_\eta(a\xi + b\zeta) = a\nabla_\eta\xi + b\nabla_\eta\zeta$,
- (iii) $\nabla_\eta(fg) = (\eta f)\xi + f\nabla_\eta\xi$.

$\nabla_\eta\xi$ được gọi là *đạo hàm thuận biến* của trường vectơ ξ dọc trường vectơ η .

Trong $\mathbb{R}^n$, đạo hàm có hướng là một liên thông affine

$$(\nabla_\eta\xi)_x = \lim_{t \rightarrow 0} \frac{\xi_{x+t\eta_x} - \xi_x}{t},$$

và được gọi là liên thông Euclide chuẩn tắc.

Mệnh đề 1.2.9. Mọi đa tạp tô pô đều có thể được trang bị một liên thông affine.

Liên thông affine có các tính chất sau:

1. Phụ thuộc vào η_x : trường vectơ $\nabla_\eta\xi$ tại điểm x phụ thuộc giá trị η_x .
2. Phụ thuộc địa phương vào ξ : trái lại, ξ_x không cung cấp đủ thông tin để xác định $\nabla_\eta\xi$ tại x . Tuy nhiên, nếu hai trường vectơ ξ và ζ trùng nhau trong một lân cận của x thì $\nabla_\eta\xi$ và $\nabla_\eta\zeta$ trùng nhau. Hơn nữa, nếu có hai liên thông affine ∇ và $\tilde{\nabla}$ thì $\nabla_\eta\xi - \tilde{\nabla}_\eta\xi$ tại x chỉ phụ thuộc giá trị ξ_x .

3. Đơn trị tại 0: giả sử ∇ và $\tilde{\nabla}$ là hai liên thông affine trên $\mathcal{M}$ và ξ, η là các trường vectơ trên $\mathcal{M}$. Khi đó, ta có

$$(\nabla_{\eta}\xi)_x = \left(\tilde{\nabla}_{\eta}\xi\right)_x \quad \text{nếu } \xi_x = 0.$$

Tính chất cuối cùng này là đặc biệt quan trọng khi xét sự hội tụ của các thuật toán tối ưu xung quanh điểm tới hạn của hàm mục tiêu.

1.2.5 Liên thông Riemann

Định nghĩa 1.2.10. Cho $\mathcal{M}$ là một đa tạp trơn với hai trường vectơ $\xi, \zeta \in C^{\infty}(T\mathcal{M})$, móc Lie của ξ và ζ tại x là hàm số

$$\begin{aligned} [\xi, \zeta]_p : C^{\infty}(\mathcal{M}) &\rightarrow \mathbb{R} \\ f &\mapsto [\xi, \zeta]_p(f) = \xi_x(\zeta(f)) - \zeta_x(\xi(f)). \end{aligned}$$

Phát biểu sau đây chỉ ra rằng móc Lie $[\xi, \zeta]_x$ là một phần tử của không gian tiếp xúc $T_x\mathcal{M}$.

Mệnh đề 1.2.11. Cho $\mathcal{M}$ là một đa tạp trơn, $\xi, \zeta \in C^{\infty}(T\mathcal{M})$, $f, g \in C^{\infty}(\mathcal{M})$ và $\lambda, \mu \in \mathbb{R}$. Khi đó

$$(i) \quad [\xi, \zeta]_x(\lambda f + \mu g) = \lambda[\xi, \zeta]_x(f) + \mu[\xi, \zeta]_x(g),$$

$$(ii) \quad [\xi, \zeta]_p(fg) = [\xi, \zeta]_x(f)g(x) + f(x)[\xi, \zeta]_x(g).$$

Định nghĩa 1.2.12. Một liên thông affine trên đa tạp $\mathcal{M}$ được gọi là *đối xứng* nếu

$$\nabla_{\eta}\xi - \nabla_{\xi}\eta = [\eta, \xi], \quad \forall \eta, \xi \in \mathfrak{X}(\mathcal{M}).$$

Định lý 1.2.13 (Levi-Civita). Trên đa tạp Riemann $\mathcal{M}$, tồn tại duy nhất một liên thông affine ∇ thỏa mãn

$$(i) \quad \nabla \text{ là liên thông affine đối xứng,}$$

(ii) $\chi\langle\eta, \xi\rangle = \langle\nabla_\chi\eta, \xi\rangle + \langle\eta, \nabla_\chi\xi\rangle, \forall \chi, \eta, \xi \in \mathfrak{X}(\mathcal{M})$ (gọi là tính tương thích với *mêtric Riemann*).

Định nghĩa 1.2.14. Liên thông affine ∇ được xác định trong Định lý Levi-Civita được gọi là *liên thông Levi-Civita* hay *liên thông Riemann* của $\mathcal{M}$.

1.2.6 Cung trắc địa, ánh xạ mũ

Ta sẽ khái quát hóa khái niệm đường thẳng trong $\mathbb{R}^n$ lên đa tạp. Nhắc lại rằng một đường thẳng trong $\mathbb{R}^n$ là ảnh của một cung tham số γ có đạo hàm cấp hai bằng 0, tức là

$$\frac{d^2}{dt^2}\gamma(t) = 0, \quad \forall t.$$

Cho đa tạp $\mathcal{M}$ với liên thông affine ∇ và $\gamma : I \rightarrow \mathcal{M}$ là một cung tham số. Kí hiệu $\mathfrak{X}(\gamma)$ là tập các trường vectơ trơn dọc cung γ .

Định nghĩa 1.2.15. Ánh xạ

$$\begin{aligned} \frac{D}{dt} : \mathfrak{X}(\gamma) &\rightarrow \mathfrak{X}(\gamma) \\ \xi &\mapsto \frac{D}{dt}\xi \end{aligned}$$

được gọi là *đạo hàm cấp 1 của trường vectơ dọc cung* γ nếu nó thỏa mãn các tính chất

- (i) $\frac{D}{dt}(a\xi + b\zeta) = a\frac{D}{dt}\xi + b\frac{D}{dt}\zeta, \quad a, b \in \mathbb{R},$
- (ii) $\frac{D}{dt}(f\xi) = f'\xi + f\frac{D}{dt}\xi, \quad f \in \mathfrak{F}(I),$
- (iii) $\frac{D}{dt}(\eta \circ \gamma)(t) = \nabla_{\dot{\gamma}(t)}\eta, \quad t \in I, \eta \in \mathfrak{X}(\mathcal{M}).$

Người ta chứng minh được đạo hàm cấp 1 của một trường vectơ dọc một cung tham số tồn tại duy nhất.

Cho $\gamma(t)$ là một cung tham số. Ánh xạ $t \mapsto \dot{\gamma}(t)$ cho ta đạo hàm của cung tham số γ .

Định nghĩa 1.2.16. Đạo hàm cấp 2 của trường vectơ dọc cung γ được xác định bởi

$$\frac{D^2}{dt^2}\gamma := \frac{D}{dt}\dot{\gamma}.$$

Chú ý rằng $\frac{D^2}{dt^2}\gamma$ phụ thuộc vào việc chọn liên thông affine trong khi $\dot{\gamma}$ thì không.

Đặc biệt, trong $\mathbb{R}^n$, với liên thông Euclide chuẩn tắc, ta có

$$\frac{D^2}{dt^2}\gamma = P_{\gamma(t)} \frac{d^2\gamma}{dt^2}. \quad (1.6)$$

Định nghĩa 1.2.17. Giả sử $\mathcal{M}$ là một đa tạp với liên thông affine ∇ . Một cung γ trên $\mathcal{M}$ được gọi là *cung trắc địa* nếu $\frac{D^2}{dt^2}\gamma(t) = 0$ với mọi t .

Chú ý rằng các liên thông affine khác nhau xác định những cung trắc địa khác nhau.

Với mọi $\xi \in T_x\mathcal{M}$, tồn tại một khoảng I chứa 0 và duy nhất một cung trắc địa $\gamma(t; x, \xi) : I \rightarrow \mathcal{M}$ sao cho $\gamma(0) = x$ và $\dot{\gamma}(0) = \xi$. Hơn nữa, ta có $\gamma(t, x, a\xi) = \gamma(at, x, \xi)$.

Định nghĩa 1.2.18. Cho đa tạp $\mathcal{M}$ với liên thông affine ∇ . Ánh xạ

$$\begin{aligned} \exp_x : T_x\mathcal{M} &\rightarrow \mathcal{M} \\ \xi &\mapsto \exp_x \xi = \gamma(1, x, \xi), \end{aligned}$$

với γ là cung trắc địa của $\mathcal{M}$, được gọi là *ánh xạ mũ* tại x .

Rộng hơn, ta có thể định nghĩa

$$\begin{aligned} \exp : T\mathcal{M} &\rightarrow \mathcal{M} \\ (x, \xi) &\mapsto \exp_x \xi = \gamma(1, x, \xi) = \gamma\left(\|\xi\|, x, \frac{\xi}{\|\xi\|}\right), \end{aligned}$$

ở đây kí hiệu (x, ξ) để chỉ vectơ ξ với gốc x . Ánh xạ $\exp$ là khả vi và $\exp_x 0_x = x, \forall x \in \mathcal{M}$.

Về mặt hình học, $\exp_x \xi$ là một điểm thuộc $\mathcal{M}$, thu được bằng cách đi chuyển một khoảng $\|\xi\|$, bắt đầu từ x , dọc theo một cung trắc địa đi qua x với vectơ vận tốc $\frac{\xi}{\|\xi\|}$.

Ví dụ 1.2.19. Xét trong $\mathbb{R}^n$. Do đạo hàm dọc một cung tham số trùng với đạo hàm thông thường nên cung trắc địa là đường thẳng. Xét đường thẳng đi qua x_0 với vectơ chỉ phương v , có phương trình $x(t) = x_0 + vt$. Khi đó, ánh xạ mũ của vectơ v tại điểm x_0 là $\exp_{x_0} v = x_0 + v$, đây chính là phép tịnh tiến điểm x_0 theo vectơ v .

Ví dụ 1.2.20. Xét mặt cầu đơn vị S^2 trong $\mathbb{R}^3$.

1. Ta chứng minh rằng một cung γ là cung trắc địa của S^2 nếu và chỉ nếu nó là đường tròn lớn của S^2 .

- Trước hết, ta có phép chiếu trực giao lên $T_x S^2$ là

$$P_x \xi = (I - xx^\top) \xi.$$

Với mỗi điểm x_0 bất kỳ của S^2 , ta hoàn toàn có thể chọn hệ tọa độ sao cho x_0 là điểm cực bắc của S^2 . Tức là $x_0 = (0, 0, 1)$. Đường tròn lớn đi qua x_0 và một điểm $(a, b, 0)$ nằm trên mặt phẳng xOy là ảnh của cung tham số

$$t \mapsto x(t) = (a \sin t, b \sin t, \cos t)^\top, \quad \text{với } a^2 + b^2 = 1.$$

Ta có

$$\begin{aligned} \frac{dx}{dt} &= (a \cos t, b \cos t, -\sin t)^\top \\ \frac{d^2x}{dt^2} &= (-a \sin t, -b \sin t, -\cos t)^\top. \end{aligned}$$

Do đó theo (1.6)

$$\begin{aligned} \frac{D^2}{dt^2} x(t) &= P_x \left(\frac{d^2x}{dt^2} \right) \\ &= (I - x(t)x(t)^\top) (-a \sin t, -b \sin t, -\cos t)^\top \\ &= - \begin{pmatrix} a^2 \sin^2 t - 1 & ab \sin^2 t & a \sin t \cos t \\ ab \sin^2 t & b^2 \sin^2 t - 1 & b \sin t \cos t \\ a \sin t \cos t & b \sin t \cos t & \cos^2 t - 1 \end{pmatrix} \begin{pmatrix} -a \sin t \\ -b \sin t \\ -\cos t \end{pmatrix} \end{aligned}$$

$$= (a^2 \sin^2 t + b^2 \sin^2 t + \cos^2 t - 1) \begin{pmatrix} a \sin t \\ b \sin t \\ \cos t \end{pmatrix} = 0.$$

Vậy mỗi đường tròn lớn của mặt cầu là một cung trắc địa của nó.

- Bây giờ xét cung trắc địa bất kỳ $\gamma(t)$ trên S^2 , lấy bất kỳ $x \in \gamma$ và ξ là một vectơ tiếp xúc của γ tại x . Gọi $E \subset \mathbb{R}^3$ là mặt phẳng đi qua tâm mặt cầu và điểm x đồng thời song song với vectơ ξ . E cắt mặt cầu theo giao tuyến là đường tròn lớn C . Kí hiệu $f : S^2 \rightarrow S^2$ là phép đối xứng qua mặt phẳng E của S^2 . Ánh xạ này là một tự đẳng cấu và nó giữ nguyên các điểm của C . Cung $f \circ \gamma$ cũng là một cung trắc địa của S^2 tại x và vectơ tiếp xúc của nó tại x cũng là ξ . Do tính duy nhất của cung trắc địa, ta phải có $f \circ \gamma = \gamma$ trong lân cận x . Điều này kéo theo phần của γ trong lân cận của x thuộc đường tròn lớn C . Do x là bất kì nên toàn bộ γ được chứa trong C . Ta có điều phải chứng minh.

2. Để xác định ánh xạ mũ, ta xét cung trắc địa đi qua điểm x_0 , theo phương vectơ ξ .

Ta có phương trình

$$\begin{cases} \|x\| = 1 \\ x = \lambda x_0 + \mu \xi \end{cases}, \quad (\lambda, \mu \in \mathbb{R}). \quad (1.7)$$

Sử dụng $\|x_0\| = 1$ và $x_0 \perp \xi$, hệ (1.7) cho ta $\lambda^2 + (\mu\|\xi\|)^2 = 1$. Vậy ta có thể đặt

$$\lambda = \cos(\alpha t), \quad \mu = \frac{1}{\|\xi\|} \sin(\alpha t),$$

khi đó

$$x(t) = x_0 \cos(\alpha t) + \frac{\xi}{\|\xi\|} \sin(\alpha t).$$

Suy ra

$$\dot{x}(t) = -x_0 \alpha \sin(\alpha t) + \frac{\xi}{\|\xi\|} \alpha \cos(\alpha t). \quad (1.8)$$

Thay $t = 0$ vào (1.8), chọn $\dot{x}(0) = \xi$, ta được $\xi = \frac{\xi}{\|\xi\|} \alpha \Rightarrow \alpha = \|\xi\|$. Ta có phương trình cung trắc địa

$$x(t) = x_0 \cos(\|\xi\|t) + \frac{\xi}{\|\xi\|} \sin(\|\xi\|t). \quad (1.9)$$

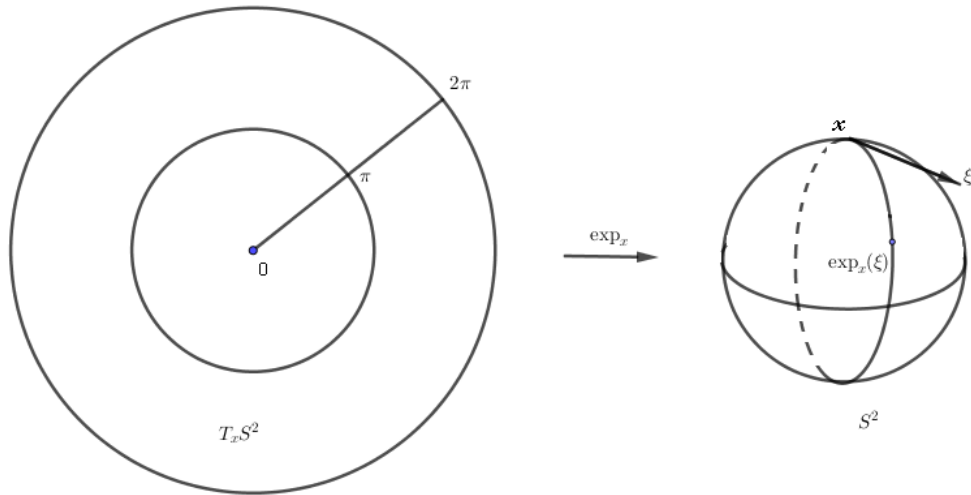
Ánh xạ mũ trong trường hợp này là

$$\exp_{x_0} \xi = x_0 \cos(\|\xi\|) + \frac{\xi}{\|\xi\|} \sin(\|\xi\|).$$

Chẳng hạn, xét các điểm $\xi \in T_{x_0} S^2$, mà $\|\xi\| = \pi$, ta có $\exp_{x_0} \xi = -x_0$. Vậy ánh xạ mũ biến đường tròn tâm 0, bán kính π trong $T_{x_0} S^2$ thành điểm xuyên tâm đối của x_0 . Với bất kì $\xi \in T_{x_0} S^2$, kí hiệu $d_g(x_0, \exp_{x_0} \xi)$ là độ dài phần cung trắc địa từ x_0 đến $\exp_{x_0} \xi$, theo (3.4), ta có:

$$d_g(x_0, \exp_{x_0} \xi) = \arccos(x_0^\top \exp_{x_0} \xi) = \|\xi\|.$$

Như vậy ta có thể hình dung, nếu x_0 là điểm cực bắc của mặt cầu thì điểm $\exp_{x_0} \xi$ được xác định bằng cách di chuyển từ cực bắc trên đường kinh tuyến theo phương ξ , một quãng bằng $\|\xi\|$.



3. Tổng quát, xét mặt cầu đơn vị S^{n-1} với mêtric Riemann và liên thông Riemann.

Cung trắc địa $t \mapsto x(t)$ được xác định như sau

$$x(t) = x(0) \cos(\|\dot{x}(0)\|t) + \dot{x}(0) \frac{1}{\|\dot{x}(0)\|} \sin(\|\dot{x}(0)\|t).$$

1.2.7 Toán tử Hessian

Ta nhắc lại ma trận Hessian của hàm giá trị thực f trên $\mathbb{R}^n$ tại điểm $x \in \mathbb{R}^n$ có phần tử ở dòng i cột j là $\partial_{ij}^2 f(x) = \frac{\partial^2 f}{\partial_i \partial_j}(x)$. Như vậy trong không gian Euclide hữu hạn chiều, về mặt hình thức, Hessian của hàm số là một ma trận. Tuy nhiên, tổng quát hơn, ta hiểu đó là một ánh xạ tuyến tính. Cái nhìn bao quát này cho phép ta định nghĩa khái niệm Hessian cho không gian Euclide bất kỳ và sau đó là cho đa tạp Riemann tổng quát.

Bây giờ ta xây dựng toán tử Hessian cho hàm f trên đa tạp. Cho f là hàm giá trị thực trên không gian Euclid $\mathcal{E}$. Toán tử Hessian tại x là toán tử tuyến tính từ $\mathcal{E}$ đến $\mathcal{E}$ xác định bởi

$$\text{Hess}f(x)[z] := \sum_{ij} \partial_{ij}^2 \hat{f}(x_1, \dots, x_n) z_j e_i,$$

với $(e_i)_{i=1, \dots, n}$ là cơ sở trực chuẩn của $\mathcal{E}$, $z = \sum_j z_j e_j$ và $\hat{f}$ là hàm số trên $\mathbb{R}^n$ xác định bởi $\hat{f}(x_1, \dots, x_n) = f(x_1 e_1 + \dots + x_n e_n)$. Dễ thấy rằng định nghĩa này không phụ thuộc vào việc chọn cơ sở trực chuẩn. Dưới đây là một phát biểu tương đương.

Định nghĩa 1.2.21. Cho f là hàm giá trị thực xác định trên không gian Euclid $\mathcal{E}$. Toán tử $\text{Hess} : \mathcal{E} \rightarrow \mathcal{E}$ được gọi là toán tử Hessian của f tại x nếu các thỏa mãn các tính chất dưới đây với mọi $y, z \in \mathcal{E}$.

- (i) $\langle \text{Hess}f(x)[y], y \rangle = D^2 f(x)[y, y] := \left. \frac{d^2}{dt^2} f(x + ty) \right|_{t=0}$,
- (ii) $\langle \text{Hess}f(x)[y], z \rangle = \langle y, \text{Hess}f(x)[z] \rangle$ (tính đối xứng).

Trong đa tạp Riemann, toán tử Hessian được khái quát hóa như sau.

Định nghĩa 1.2.22. Cho f là hàm giá trị thực xác định trên đa tạp Riemann $\mathcal{M}$, *Hessian Riemann của f tại $x \in \mathcal{M}$* là ánh xạ tuyến tính $\text{Hess}f(x) : T_x \mathcal{M} \rightarrow T_x \mathcal{M}$ xác định bởi

$$\text{Hess}f(x)[\xi_x] = \nabla_{\xi_x} \text{grad}f, \quad \forall \xi_x \in T_x \mathcal{M},$$

với ∇ là liên thông Riemann trên $\mathcal{M}$.

Hessian Riemann của f có các tính chất sau

Mệnh đề 1.2.23. *Giả sử f là hàm giá trị thực xác định trên đa tạp Riemann $\mathcal{M}$. Khi đó, ta có*

$$(i) \quad \langle \text{Hess}f[\xi], \eta \rangle = \xi(\eta f) - (\nabla_\xi \eta)f, \quad \forall \xi, \eta \in \mathfrak{X}(\mathcal{M}),$$

$$(ii) \quad \text{Hess}f \text{ là đối xứng, tức là } \langle \text{Hess}f[\xi], \eta \rangle = \langle \xi, \text{Hess}f[\eta] \rangle, \quad \forall \xi, \eta \in \mathfrak{X}(\mathcal{M}).$$

Kết quả dưới đây cho thấy Hessian Riemann của f tại x trùng với toán tử Hessian của $f \circ \exp_x$ tại $0_x \in T_x\mathcal{M}$. Chú ý rằng $f \circ \exp_x$ là hàm giá trị thực trên không gian tiếp xúc $T_x\mathcal{M}$.

Mệnh đề 1.2.24. *Giả sử $\mathcal{M}$ là một đa tạp Riemann và f là hàm giá trị thực trên $\mathcal{M}$. Khi đó*

$$\text{Hess}f(x) = \text{Hess}(f \circ \exp_x)(0_x), \quad \forall x \in \mathcal{M},$$

với $\text{Hess}f(x)$ là Hessian Riemann của $f : \mathcal{M} \rightarrow \mathbb{R}$ tại x và $\text{Hess}(f \circ \exp_x)(0_x)$ là toán tử Hessian của hàm $f \circ \exp_x$ tại $0_x \in T_x\mathcal{M}$.

Chương 2

Thuật toán Tìm theo đường thẳng trên đa tạp

Trong chương này, chúng tôi sẽ trình bày thuật toán Tìm theo đường thẳng (*line search*) để tối ưu hóa hàm số xác định trên đa tạp Riemann. Trước tiên, chúng tôi nhắc lại sơ lược thuật toán này cho hàm số xác định trên $\mathbb{R}^n$. Sau đó, chúng tôi giải thích tại sao khi áp dụng thuật toán này cho hàm trên đa tạp, ta cần phải sử dụng những công cụ lý thuyết của đa tạp. Khi trình bày, chúng tôi chủ yếu dựa vào 2 tài liệu [5], [6].

2.1 Thuật toán Tìm theo đường thẳng trong $\mathbb{R}^n$

Thuật toán Tìm theo đường thẳng là một trong những chiến lược cơ bản để tìm điểm cực tiểu x^* của hàm mục tiêu $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Với x_k đã cho, sơ đồ lặp là

$$x_{k+1} = x_k + t_k \eta_k.$$

Ta cần tìm hướng η_k mà dọc theo đó hàm mục tiêu f sẽ bị giảm và sau đó tính toán cỡ bước t_k xác định khoảng cách sẽ di chuyển dọc theo hướng đó. Hướng η_k có thể được tính bằng các phương pháp: gradient, phương pháp Newton, phương pháp tựa Newton. Cỡ bước có thể được xác định chính xác hoặc không chính xác.

Người ta thường phải tìm cực tiểu x^* của các hàm mục tiêu $f : \mathbb{R}^n \rightarrow \mathbb{R}$ thỏa mãn giả thiết tiêu chuẩn.

Định nghĩa 2.1.1. Hàm mục tiêu $f : \mathbb{R}^n \rightarrow \mathbb{R}$ được gọi là thỏa mãn *giả thiết tiêu chuẩn* tại x^* nếu nó thỏa mãn các điều kiện sau

(i) f khả vi cấp hai và thỏa mãn điều kiện *Lipschitz* đối với *Hessian*

$$\|\text{Hess}f(x) - \text{Hess}f(y)\| \leq \gamma\|x - y\|,$$

ta cũng nói f có đạo hàm liên tục *Lipschitz* hai lần với hằng số γ ;

(ii) $\text{grad}f(x^*) = 0$;

(iii) $\text{Hess}f(x^*)$ là xác định dương.

Định nghĩa 2.1.2. Giả sử $\{x_k\}$ là một dãy của $\mathbb{R}^n$.

1. Dãy $\{x_k\}$ được gọi là *hội tụ q - bậc hai* tới $x^* \in \mathbb{R}^n$ nếu x_k hội tụ tới x^* và tồn tại $K > 0$ sao cho

$$\|x_{k+1} - x^*\| \leq K\|x_k - x^*\|^2, \quad \forall k.$$

2. Dãy $\{x_k\}$ được gọi là *hội tụ q - siêu tuyến tính* tới $x^* \in \mathbb{R}^n$ nếu

$$\lim_{k \rightarrow \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = 0.$$

3. Dãy $\{x_k\}$ được gọi là *hội tụ tuyến tính* tới $x^* \in \mathbb{R}^n$ nếu tồn tại $c \in (0, 1)$ sao cho

$$\|x_{k+1} - x^*\| \leq c\|x_k - x^*\|, \quad \forall k.$$

Định nghĩa 2.1.3. Một phương pháp lặp tính x^* được gọi là hội tụ địa phương q - (bậc hai, siêu tuyến tính, tuyến tính) nếu dãy lặp $\{x_k\}$ sinh bởi phương pháp đó hội tụ q - (bậc hai, siêu tuyến tính, tuyến tính) tới x^* , miễn là x_0 đủ gần x^* .

2.1.1 Phương pháp giảm sâu nhất

Định nghĩa 2.1.4. Vectơ $d \in \mathbb{R}^n$ được gọi là một *hướng giảm* của f tại x nếu

$$\left. \frac{df(x + t\eta)}{dt} \right|_{t=0} = \text{grad}f(x)^\top \eta < 0.$$

Để thấy $-\text{grad}f(x)$ là một hướng giảm. Tổng quát hơn, bất kỳ hướng nào tạo với $\text{grad}f(x)$ một góc tù đều là hướng giảm.

Hướng giảm sâu nhất tại x là $\eta = -\text{grad}f(x)$. Phương pháp này cập nhật bởi công thức

$$x_{k+1} = x_k - t\text{grad}f(x_k),$$

trong đó $t > 0$ là cỡ bước. Nếu ta chọn đơn giản $t = 1$ thì không đảm bảo được x_{k+1} sẽ gần x^* hơn x_k , ngay cả khi x_k là rất gần x^* và thỏa mãn giả thiết tiêu chuẩn. Để làm cho phương pháp giảm sâu nhất thành công, ta cần chọn độ dài bước λ phù hợp. Tài liệu [6] đã đưa ra cách chọn $t = \beta^m$, với $\beta \in (0, 1)$ và m là số nguyên không âm nhỏ nhất sao cho

$$f(x_k - t\text{grad}f(x_k)) - f(x_k) < -\alpha t \|\text{grad}f(x_k)\|^2, \quad \alpha \in (0, 1). \quad (2.1)$$

Điều kiện (2.1) chính là điều kiện giảm đủ hay điều kiện Armijo áp dụng cho phương pháp giảm sâu nhất. Ta có thuật toán Tìm theo đường thẳng với phương pháp giảm sâu nhất.

Algorithm 1 Phương pháp giảm sâu nhất

Require: Hàm f thỏa mãn giả thiết tiêu chuẩn.

Input: Giá trị ban đầu x_0, τ_r, τ_a , hàm f .

Output: Dãy $\{x_k\}$ hội tụ đến điểm tối hạn của f .

- 1: Đặt $r_0 = \|\text{grad}f(x_0)\|$.
 - 2: **for** $k = 0, 1, 2, \dots$ **do**
 - 3: Tính $f(x_k)$, $\text{grad}f(x_k)$ và kiểm tra điều kiện dừng $\|\text{grad}f(x_k)\| \leq \tau_r r_0 + \tau_a$.
 - 4: Tìm số nguyên dương m nhỏ nhất sao cho (2.1) thỏa mãn với $t = \beta^m, \beta \in (0, 1)$.
 - 5: Đặt $x_{k+1} = x_k - t\text{grad}f(x_k)$.
 - 6: **end for**
-

2.1.2 Phương pháp Newton

Xét bài toán tìm cực tiểu x^* của hàm mục tiêu $f : \mathbb{R}^n \rightarrow \mathbb{R}$ thỏa mãn giả thiết tiêu chuẩn. Phương pháp Newton chọn $x_0 \in \mathbb{R}^n$ và dựa trên cập nhật

$$x_{k+1} = x_k - (\text{Hess}f(x_k))^{-1}(\text{grad}f(x_k)). \quad (2.2)$$

Phương trình (2.2) được gọi là phương trình Newton. Do tính xác định dương của $\text{Hess}f(x_k)$, hướng $-(\text{Hess}f(x_k))^{-1}\text{grad}f(x_k)$ cũng là một hướng giảm.

Ta có thuật toán

Algorithm 2 Phương pháp Newton

Input: x_0, f, τ_r, τ_a

Output: Dãy $\{x_k\}$.

- 1: Đặt $r = \|\text{grad}f(x_0)\|$
 - 2: **while** $\|\text{grad}f(x_k)\| > \tau_r r + \tau_a$ **do**
 - 3: Tính $\text{Hess}f(x_k)$ và giải hệ phương trình $\text{Hess}f(x_k)s = -\text{grad}f(x_k)$
 - 4: Đặt $x_{k+1} = x_k + s$
 - 5: **end while**
-

Về sự hội tụ của phương pháp Newton, ta có định lý sau.

Định lý 2.1.5. *Giả sử giả thiết tiêu chuẩn được thỏa mãn. Khi đó tồn tại $\delta > 0$ sao cho nếu $x_k \in B(x^*, \delta)$, bước lặp (2.2) hội tụ q - bậc hai đến x^* .*

2.2 Tìm theo đường thẳng trên đa tạp Riemann tổng quát

2.2.1 Phân tích

Như đã trình bày ở Mục 2.1, khi cần tối ưu hóa hàm mục tiêu khả vi $f(x)$ trong $\mathbb{R}^n$, người ta tính tiếp liên tiếp điểm thử nghiệm x theo hướng giảm sâu nhất, $\eta := -\text{grad}f(x)$, cho đến khi tìm được điểm x_* mà gradient triệt tiêu. Điểm x^* đó được gọi là *điểm tối hạn* của f .

Trong $\mathbb{R}^n$, việc tịnh tiến theo hướng của một vectơ rất rõ ràng: ta chỉ cần cộng vào tọa độ của điểm cần tịnh tiến với tọa độ vectơ tịnh tiến. Nhưng trên một đa tạp, gradient của hàm mục tiêu lại là một vectơ thuộc không gian tiếp xúc của đa tạp. Khi đó, về nguyên tắc, ta không thể cộng một điểm trên đa tạp với một vectơ trên không gian tiếp xúc. Nếu giả sử có cộng được một cách máy móc thì kết quả đó sẽ không còn là một điểm nằm trên đa tạp nữa.

Vì vậy đối với đa tạp, ta cần xét sự dịch chuyển theo hướng của vectơ tiếp xúc thông qua khái niệm ánh xạ rút.

Định nghĩa 2.2.1 (Ánh xạ rút). Cho đa tạp $\mathcal{M}$. Với mỗi $x \in \mathcal{M}$, ánh xạ $R_x : T_x\mathcal{M} \rightarrow \mathcal{M}$ được gọi là ánh xạ rút tại x nếu các điều kiện sau được thỏa mãn:

$$(i) \ R_x(0_x) = x;$$

(ii) Nếu đồng nhất $T_{0_x}T_x\mathcal{M} \simeq T_x\mathcal{M}$, ta có

$$DR_x(0_x) = \text{id}_{T_x\mathcal{M}}. \quad (2.3)$$

Với mọi vectơ tiếp xúc $\xi \in T_x\mathcal{M}$, cung tham số $\gamma_\xi : t \mapsto R_x(t\xi)$ thỏa mãn $\dot{\gamma}_\xi(0) = \xi$. Như vậy, thay vì tịnh tiến theo vectơ ξ để được $x + \xi \notin \mathcal{M}$ ta sử dụng $R_x(\xi) \in \mathcal{M}$ mà vẫn không làm thay đổi hướng tìm kiếm ξ . Ánh xạ rút đã khắc phục được khó khăn nêu ra ở trên.

Một mục đích quan trọng nữa của ánh xạ rút là nó giúp chuyển hàm mục tiêu được định nghĩa trong lân cận điểm $x \in \mathcal{M}$ thành hàm mục tiêu định nghĩa trong không gian vectơ $T_x\mathcal{M}$. Đặc biệt, cho trước một hàm giá trị thực f trên $\mathcal{M}$ được trang bị một ánh xạ rút R_x tại x , hàm số

$$\hat{f}_x = f \circ R_x \quad (2.4)$$

được gọi là *nói dài* của f dọc theo R_x . Chú ý rằng với điều kiện (2.3), ta có $D\hat{f}_x(0_x) = Df(x)$. Nếu $\mathcal{M}$ được trang bị một mêtric Riemann, ta có:

$$\text{grad}\hat{f}_x(0_x) = \text{grad}f(x). \quad (2.5)$$

Ví dụ 2.2.2. Xét mặt cầu S^{n-1} trong $\mathbb{R}^n$. Với mỗi $x \in S^{n-1}$ ta có ánh xạ rút

$$R_x : T_x S^{n-1} \rightarrow S^{n-1}$$

$$\xi \mapsto R_x(\xi) = \frac{x + \xi}{\|x + \xi\|}.$$

Chứng minh.

(i) $R_x(0_x) = \frac{x + 0_x}{\|x + 0_x\|} = x$, vì $\|x\| = 1$.

(ii) Ta cần chứng minh $DR_x(0_x)[\xi] = \xi, \forall \xi \in T_x S^{n-1}$. Đặt $\mathcal{N} = \{x \in \mathbb{R} : x > 0\}$.

Xét ánh xạ

$$\phi : S^{n-1} \times \mathcal{N} \rightarrow \mathbb{R}_*^n$$

$$(x, r) \mapsto rx$$

và phép chiếu lên thành phần thứ nhất

$$\pi_1 : S^{n-1} \times \mathcal{N} \rightarrow S^{n-1}$$

$$(x, y) \mapsto x.$$

Dễ thấy ϕ là một vi phôi và $R_x(\xi) = (\pi_1 \circ \phi^{-1})(x + \xi)$. Ta có

$$DR_x(0_x)[\xi] = D(\pi_1 \circ \phi^{-1})(x)[\xi] = D\pi_1(\phi^{-1}(x)) [D\phi^{-1}(x)[\xi]]. \quad (2.6)$$

Áp dụng (1.2), ta được

$$D\phi^{-1}(x)[\xi] = D\phi^{-1}(\phi(x, 1))[\xi] = (D\phi(x, 1))^{-1} [\xi]. \quad (2.7)$$

Mặt khác, áp dụng (1.1), ta có

$$D\phi(x, 1)[\xi] = \lim_{t \rightarrow 0} \frac{\phi(x + t\xi, 1) - \phi(x, 1)}{t} = \lim_{t \rightarrow 0} \frac{x + t\xi - x}{t} = \xi.$$

Nên (2.7) cho ta $D\phi^{-1}(x)[\xi] = \xi$. Thay vào (2.6),

$$DR_x(0_x)[\xi] = D\pi_1(\phi^{-1}(x)) [\xi] = D\pi_1(x, 1)[\xi] = \xi.$$

Ta có điều phải chứng minh. □

Về mặt hình học, $R_x(\xi)$ là điểm thuộc S^{n-1} có khoảng cách đến $x + \xi$ nhỏ nhất.

Khẳng định sau đây chỉ ra, ánh xạ mũ là một trường hợp riêng của ánh xạ rút.

Mệnh đề 2.2.3. *Cho đa tạp $\mathcal{M}$ với liên thông affine ∇ . Ánh xạ mũ trên đa tạp $\mathcal{M}$ được sinh bởi ∇ là một ánh xạ rút.*

Mệnh đề 2.2.4. *Giả sử R là một ánh xạ rút thỏa mãn*

$$\left. \frac{D^2}{dt^2} R(t\xi) \right|_{t=0} = 0, \quad \forall \xi \in T_x \mathcal{M}, \quad (2.8)$$

với $\frac{D^2}{dt^2} \gamma$ là đạo hàm cấp 2 dọc cung γ . Khi đó

$$\text{Hess} f(x) = \text{Hess}(f \circ R_x)(0_x).$$

Ý nghĩa của Mệnh đề 2.2.4 là ở việc thay thế hàm mũ $\exp$ ở Mệnh đề 1.2.24, vốn rất khó để tính toán trong hầu hết trường hợp, bằng ánh xạ rút thỏa mãn điều kiện (2.8) vốn dễ tính toán.

Mệnh đề 2.2.5. *Giả sử R là một ánh xạ rút và x là một điểm tới hạn của hàm giá trị thực f , tức là $\text{grad} f(x) = 0$. Khi đó*

$$\text{Hess} f(x) = \text{Hess}(f \circ R_x)(0_x).$$

Như vậy, so với Mệnh đề 2.2.4, điều kiện (2.8) không cần dùng đến nếu x là một điểm tới hạn của hàm f .

2.2.2 Thuật toán Tìm theo đường thẳng

Như đã chỉ ra ở phần trước, lược đồ lặp theo phương pháp giảm sâu không thực hiện được trên đa tạp. Ý tưởng để khắc phục việc này là thay thế lặp $x_k - t_k \eta_k$ bởi $R_{x_k}(t_k \eta_k)$, tức là lấy x_k làm gốc và "rút" vectơ $t_k \eta_k$ về thành điểm x_{k+1} . Nói một

cách khác, ta di chuyển điểm x_k theo cung $R_{x_k}(t_k\eta_k)$ tạo bởi ánh xạ rút R_{x_k} . Theo đó, phương pháp Tìm theo đường thẳng trên đa tạp dựa trên cập nhật

$$x_{k+1} = R_{x_k}(t_k\eta_k)$$

với $\eta_k \in T_{x_k}\mathcal{M}$ và t_k là một vô hướng. Với mỗi ánh xạ rút R được chọn, ta cần chọn hướng tìm kiếm η_k và cỡ bước t_k sao cho thu được kết quả hội tụ toàn cục.

Định nghĩa 2.2.6. Cho trước hàm mục tiêu f trên đa tạp Riemann $\mathcal{M}$, một dãy $\{\eta_k\}$, $\eta_k \in T_{x_k}\mathcal{M}$ được gọi là *dãy liên kết gradient* nếu với bất kì dãy con $\{x_k\}_{k \in K}$ của $\{x_k\}$ hội tụ tới một điểm tới hạn của f , dãy con tương ứng $\{\eta_k\}_{k \in K}$ là giới nội và

$$\limsup_{k \rightarrow \infty, k \in K} \langle \text{grad} f(x_k), \eta_k \rangle < 0.$$

Định nghĩa 2.2.7. Cho trước hàm mục tiêu f trên đa tạp Riemann $\mathcal{M}$ với ánh xạ rút R , điểm $x \in \mathcal{M}$, vectơ tiếp tuyến $\eta \in T_x\mathcal{M}$ và $\alpha > 0, \beta, \sigma \in (0, 1)$, *điểm Armijo* là $\eta^A = t^A\eta = \beta^m\alpha\eta$, với số nguyên m không âm nhỏ nhất thỏa mãn

$$f(x) - f(R_x(\beta^m\alpha\eta)) \geq -\sigma \langle \text{grad} f(x), \beta^m\alpha\eta \rangle_x. \quad (2.9)$$

Số thực t^A được gọi là *cỡ bước Armijo*.

Lấy hướng giảm là $-\sigma \text{grad} f(x)$, Điều kiện 2.9 cho ta bất đẳng thức 2.2 quen thuộc.

Ta có Thuật toán 3.

Nếu có một phương pháp tính toán hiệu quả để tối ưu hóa $f \circ R_{x_k}$ trong không gian con 2 chiều của $T_{x_k}\mathcal{M}$ thì một lựa chọn khả thi cho x_{k+1} trong Bước 3 là $R_{x_k}(\xi_k)$, với ξ_k được định nghĩa bởi:

$$\xi_k := \arg \min_{\xi \in S_k} f(R_{x_k}(\xi)), \quad S_k := \text{span}\{\eta_k, R_{x_k}^{-1}(x_{k-1})\},$$

ở đây $\text{span}\{u, v\} = \{au + bv : a, b \in \mathbb{R}\}$. Đây là một tối ưu hóa trên một không gian con 2 chiều S_k của $T_{x_k}\mathcal{M}$. Rõ ràng S_k chứa điểm Armijo tương ứng với η_k , do đó $\eta_k \in S_k$. Điều đó kéo theo (2.10) với $c = 1$.

Algorithm 3 Tìm theo đường thẳng

Require: Đa tạp Riemann $\mathcal{M}$, hàm f khả vi liên tục trên $\mathcal{M}$; ánh xạ rút R từ $T\mathcal{M}$ vào $\mathcal{M}$, vô hướng $\alpha > 0$, $c, \beta, \sigma \in (0, 1)$.

Input: Giá trị ban đầu $x_0 \in \mathcal{M}$.

Output: Dãy $\{x_k\}$.

1: **for** $k = 0, 1, 2, \dots$ **do**

2: Chọn $\eta_k \in T_{x_k}\mathcal{M}$ sao cho dãy $\{\eta_i\}_{i=0,1,\dots}$ là dãy liên kết gradient.

3: Chọn x_{k+1} sao cho:

$$f(x_k) - f(x_{k+1}) \geq c(f(x_k) - f(R_{x_k}(t_k^A \eta_k))), \quad (2.10)$$

trong đó t_k^A là cỡ bước Armijo với $\alpha, \beta, \sigma, \eta_k$ cho trước.

4: **end for**

2.2.3 Sự hội tụ của thuật toán Tìm theo đường thẳng

Định nghĩa 2.2.8. Dãy $\{x_k\}_{k=0,1,\dots}$ các điểm của đa tạp $\mathcal{M}$ được gọi là *hội tụ* nếu tồn tại một bản đồ (U, φ) của $\mathcal{M}$, một điểm $x^* \in U$, và một số $K > 0$ sao cho $x_k \in U, \forall k > K$ và dãy $\{\varphi(x_k)\}_{k=K,K+1,\dots}$ hội tụ về $\varphi(x^*)$.

Điểm $\varphi^{-1}(\lim_{k \rightarrow \infty} \varphi(x_k))$ khi đó được gọi là *giới hạn* của dãy $\{x_k\}_{k=0,1,\dots}$.

Dễ thấy giới hạn của một dãy hội tụ trên đa tạp là duy nhất.

Định nghĩa 2.2.9. Cho dãy $\{x_k\}_{k=0,1,\dots}$, điểm x được gọi là một *điểm giới hạn* của dãy $\{x_k\}_{k=0,1,\dots}$ nếu tồn tại một dãy con $\{x_{j_k}\}_{k=0,1,\dots}$ hội tụ về x .

Tập hợp các điểm giới hạn của một dãy được gọi là *tập giới hạn* của dãy đó.

Định lý 2.2.10. Giả sử $\{x_k\}$ là dãy lặp xác định trong Thuật toán 3. Khi đó mọi điểm giới hạn của $\{x_k\}$ đều là điểm tới hạn của hàm mục tiêu f .

Chứng minh. Ta chứng minh định lý bằng phản chứng. Giả sử rằng có một dãy con $\{x_k\}_{k \in \mathcal{K}}$ hội tụ về x^* mà $\text{grad} f(x^*) \neq 0$. Vì dãy $\{f(x_k)\}$ không tăng nên dãy $\{f(x_k)\}$ hội tụ tới $f(x^*)$. Do đó $f(x_k) - f(x_{k+1})$ hội tụ về 0. Dựa vào thuật toán,

$$f(x_k) - f(x_{k+1}) \geq -c\sigma\alpha_k \langle \text{grad} f(x_k), \eta_k \rangle_{x_k}.$$

Do $\{\eta_k\}$ là dãy liên kết gradient, ta phải có $\{\alpha_k\}_{k \in \mathcal{K}} \rightarrow 0$. Các số α_k được xác định từ quy tắc Armijo, vì vậy, với mọi số $k > \bar{k}$, $\alpha_k = \beta^{m_k} \alpha$, ở đây m_k là một số nguyên lớn hơn hoặc bằng 0. Điều này có nghĩa là cập nhật $\frac{\alpha_k}{\beta \eta_k}$ không thỏa mãn điều kiện Armijo. Do đó

$$f(x_k) - f\left(R_{x_k}\left(\frac{\alpha_k}{\beta} \eta_k\right)\right) < -\sigma \frac{\alpha_k}{\beta} \langle \text{grad} f(x_k), \eta_k \rangle_{x_k}, \quad \forall k \in \mathcal{K}, k \geq \bar{k}.$$

Đặt

$$\tilde{\eta}_k = \frac{\eta_k}{\|\eta_k\|}, \quad \tilde{\alpha}_k = \frac{\alpha_k \|\eta_k\|}{\beta}.$$

bất đẳng thức trên trở thành

$$\frac{\hat{f}_{x_k}(0) - \hat{f}_{x_k}(\tilde{\alpha}_k \tilde{\eta}_k)}{\tilde{\alpha}_k} < -\sigma \langle \text{grad} f(x_k), \tilde{\eta}_k \rangle_{x_k}, \quad \forall k \in \mathcal{K}, k \geq \bar{k},$$

ở đây $\hat{f}$ được định nghĩa ở (2.4). Theo định lý giá trị trung bình, tồn tại $t \in [0, \tilde{\alpha}_k]$ sao cho

$$-D\hat{f}_{x_k}(t\tilde{\eta}_k)[\tilde{\eta}_k] < -\sigma \langle \text{grad} f(x_k), \tilde{\eta}_k \rangle_{x_k}, \quad \forall k \in \mathcal{K}, k \geq \bar{k}. \quad (2.11)$$

Vì $\{\alpha_k\}_{k \in \mathcal{K}} \rightarrow 0$ và η_k là dãy liên kết gradient nên $\{\tilde{\alpha}_k\}_{k \in \mathcal{K}} \rightarrow 0$. Hơn nữa, vì $\tilde{\eta}_k$ có chuẩn bằng 1 nên nó thuộc một tập compact, do đó tồn tại một tập chỉ số $\tilde{\mathcal{K}} \subseteq \mathcal{K}$ sao cho $\{\tilde{\eta}_k\}_{k \in \tilde{\mathcal{K}}} \rightarrow \tilde{\eta}^*$ với $\|\tilde{\eta}^*\| = 1$. Bây giờ ta chuyển qua giới hạn trong (2.11) trên $\tilde{\mathcal{K}}$. Do mêtric Riemann là liên tục và $f \in C^1$, $D\hat{f}_{x_k}(0)[\tilde{\eta}_k] = \langle \text{grad} f(x_k), \tilde{\eta}_k \rangle_{x_k}$, ta có

$$-\langle \text{grad} f(x^*), \tilde{\eta}^* \rangle_{x^*} \leq -\sigma \langle \text{grad} f(x^*), \tilde{\eta}^* \rangle_{x^*}.$$

Do $\sigma < 1$ nên $\langle \text{grad} f(x^*), \tilde{\eta}^* \rangle_{x^*} \geq 0$. Nhưng $\{\eta_k\}$ là dãy liên kết gradient nên

$$\langle \text{grad} f(x^*), \tilde{\eta}^* \rangle_{x^*} < 0.$$

Mâu thuẫn này cho ta khẳng định của mệnh đề. □

Hệ quả 2.2.11. Giả sử $\{x_k\}$ là dãy lặp được xây dựng trong Thuật toán 3 và tập $\mathcal{L} = \{x \in M : f(x) \leq f(x_0)\}$ compact. Khi đó $\lim_{k \rightarrow \infty} \|\text{grad} f(x_k)\| = 0$.

Chứng minh. Ta chứng minh bằng phản chứng. Giả sử ngược lại, khi đó tồn tại một dãy con $\{x_k\}_{k \in \mathcal{K}}$ và $\varepsilon > 0$ sao cho

$$\|\text{grad} f(x_k)\| > \varepsilon, \quad \forall k \in \mathcal{K}.$$

Vì f không tăng trên $\{x_k\}$ nên $x_k \in \mathcal{L}$ với mọi k . Do $\mathcal{L}$ là compact nên dãy $\{x_k\}_{k \in \mathcal{K}}$ có một điểm giới hạn $x^* \in \mathcal{L}$. Từ tính chất liên tục của $\text{grad} f$, có $\|\text{grad} f(x^*)\| \geq \varepsilon$. Tức là x^* không phải là một điểm tới hạn. Mâu thuẫn với Định lý 2.2.10. $\square$

2.2.4 Tốc độ hội tụ

Định lý dưới đây cho ta cơ sở tính toán kết quả hội tụ địa phương của các phương pháp lặp xác định bởi $x_{k+1} = F(x_k)$ với ánh xạ trơn $F : \mathcal{M} \rightarrow \mathcal{M}$.

Định lý 2.2.12. *Giả sử $F : \mathcal{M} \rightarrow \mathcal{M}$ là ánh xạ lớp C^1 mà cả tập nguồn và tập đích đều chứa một lân cận của x^* .*

(i) *Nếu $DF(x^*) = 0$ thì phương pháp lặp với ánh xạ lặp F là hội tụ siêu tuyến tính địa phương tới x^* .*

(ii) *Nếu $DF(x^*) = 0$ và F là ánh xạ lớp C^2 thì phương pháp lặp với ánh xạ lặp F là hội tụ bậc hai địa phương tới x^* .*

Định lý 2.2.13. *Giả sử $\{x_k\}$ là dãy lặp xác định bởi Thuật toán 3 với $\eta_k = -\text{grad} f(x_k)$, hội tụ tới điểm x^* . Đặt $\lambda_{H,\min}$, và $\lambda_{H,\max}$ lần lượt là giá trị riêng nhỏ nhất và lớn nhất của ma trận Hessian của f tại x^* . Giả sử rằng $\lambda_{H,\min} > 0$ (do đó, x^* là một điểm cực tiểu địa phương của f). Khi đó, cho trước r thuộc khoảng $(r^*, 1)$, với $r^* = 1 - \min \left\{ 2\sigma\alpha\lambda_{H,\min}, 4\sigma(1-\sigma)\beta \frac{\lambda_{H,\min}}{\lambda_{H,\max}} \right\}$, tồn tại một số nguyên $K \geq 0$ sao cho*

$$f(x_{k+1}) - f(x^*) \leq (r + (1-r)(1-c))(f(x_k) - f(x^*))$$

với mọi $k \geq K$, ở đây c là tham số trong Thuật toán 3.

2.3 Phương pháp Newton

Cho đa tạp Riemann $\mathcal{M}$ được trang bị ánh xạ rút R . Chiến lược của phương pháp Newton là tìm x_{k+1} từ x_k theo cách sau:

1. Tìm một vectơ tiếp xúc $\eta_k \in T_{x_k}\mathcal{M}$ là nghiệm của phương trình Newton.
2. Dùng ánh xạ rút để đưa η_k về x_{k+1} .

2.3.1 Phương pháp Newton trên đa tạp Riemann với hàm mục tiêu giá trị thực

Bây giờ ta xét f là hàm mục tiêu giá trị thực trên đa tạp Riemann $\mathcal{M}$. Phương trình Newton

$$\text{Hess}f(x_k)\eta_k = -\text{grad}f(x_k), \quad (2.12)$$

với

$$\begin{aligned} \text{Hess}f(x) : T_x\mathcal{M} &\rightarrow T_x\mathcal{M} \\ \eta &\mapsto \nabla_\eta \text{grad}f \end{aligned} \quad (2.13)$$

là Hessian của f tại x với liên thông affine ∇ . Ta có Thuật toán 4.

Trong trường hợp tổng quát, vectơ Newton η_k , nghiệm của phương trình (2.12), không nhất thiết là một hướng giảm của f . Thật vậy, ta có:

$$Df(x_k)[\eta_k] = \langle \text{grad}f(x_k), \eta_k \rangle = - \langle \text{grad}f(x_k), (\text{Hess}f(x_k))^{-1} \text{grad}f(x_k) \rangle, \quad (2.14)$$

không chắc chắn âm nếu không bổ sung điều kiện cho toán tử $\text{Hess}f(x_k)$. Một điều kiện đủ để η_k là hướng giảm là $\text{Hess}f(x_k)$ xác định dương, tức là

$$\langle \xi, \text{Hess}f(x_k)[\xi] \rangle > 0, \quad \forall \xi \neq 0.$$

Khi ∇ là một liên thông affine đối xứng (như liên thông Riemann chẳng hạn) thì $\text{Hess}f(x_k)$ là xác định dương khi và chỉ khi tất cả các giá trị riêng của nó đều dương.

Algorithm 4 Phương pháp Newton trên đa tạp Riemann với hàm mục tiêu giá trị thực

Require: Đa tạp Riemann $\mathcal{M}$, ánh xạ rút R trên $\mathcal{M}$, liên thông affine ∇ trên $\mathcal{M}$, hàm mục tiêu f giá trị thực trên $\mathcal{M}$.

Goal: : Tìm một điểm tối hạn của f , tức là tìm $x \in \mathcal{M}$ sao cho $\text{grad}f(x) = 0$.

Input: Giá trị ban đầu $x_0 \in \mathcal{M}$.

Output: Dãy $\{x_k\}$.

1: **for** $k = 0, 1, 2, \dots$ **do**

2: Giải phương trình Newton

$$\text{Hess}f(x_k)\eta_k = -\text{grad}f(x_k)$$

với ẩn $\eta_k \in T_{x_k}M$, ở đây $\text{Hess}f(x_k)\eta_k := \nabla_{\eta_k}\text{grad}f$.

3: Đặt $x_{k+1} := R_x(\eta_k)$.

4: **end for**

Trong thực hành, để có được kết quả hội tụ, ta sử dụng phương pháp bán Newton, chọn vectơ η_k là nghiệm của

$$(\text{Hess}f(x_k) + E_k)\eta_k = -\text{grad}f(x_k),$$

ở đây toán tử E_k được chọn sao cho $(\text{Hess}f(x_k) + E_k)$ là xác định dương. Với E_k được lựa chọn phù hợp, dãy $\{\eta_k\}$ là dãy liên kết gradient, do đó giả thiết về sự hội tụ toàn cục trong Thuật toán 3 được thỏa mãn (Định lý 2.2.10). Cần lưu ý rằng E_k phải đảm bảo không phá vỡ cấu trúc hội tụ siêu tuyến tính của dãy Newton thuần túy khi hội tụ về điểm tối hạn.

2.3.2 Sự hội tụ địa phương

Bổ đề 2.3.1. Giả sử $\|\cdot\|$ là một chuẩn bất kỳ trên $\mathbb{R}^{n \times n}$ sao cho $\|I\| = 1$. Nếu $\|E\| < 1$ thì $(I - E)^{-1}$ tồn tại và

$$\|(I - E)^{-1}\| \leq \frac{1}{1 - \|E\|}.$$

Nếu A không suy biến và $\|A^{-1}(B - A)\| < 1$, khi đó B không suy biến và

$$\|B^{-1}\| \leq \frac{\|A^{-1}\|}{1 - \|A^{-1}(B - A)\|}.$$

Định lí 2.3.2 (Sự hội tụ địa phương của phương pháp Newton). Với các điều kiện trong Thuật toán 4, giả sử tồn tại $x^* \in \mathcal{M}$ sao cho $\xi_{x^*} = 0$ và $(\text{Hess}f(x^*))^{-1}$ tồn tại. Khi đó tồn tại một lân cận U của x^* trong $\mathcal{M}$ sao cho với mọi $x_0 \in U$, Thuật toán 4 xây dựng được một dãy $\{x_k\}$ hội tụ siêu tuyến tính (ít nhất là hội tụ bậc hai) về x^* .

Chương 3

Ví dụ về bài toán tối ưu trên mặt cầu

Trong chương này, chúng tôi sẽ sử dụng thuật toán Newton, để giải quyết một số bài toán tối ưu cụ thể trên mặt cầu, đó là bài toán K-mean, bài toán điểm trung chuyển hàng không, bài toán giá trị riêng. Các ví dụ số được chúng tôi thực hành trên MATLAB. Nội dung chính của 3.3 có thể được tìm thấy ở tài liệu [5].

3.1 Bài toán K-mean trên mặt cầu

3.1.1 Bài toán

Định nghĩa 3.1.1. Trong không gian mêtric (M, d) , cho các điểm $x_1, \dots, x_n$. Điểm $x^* \in M$ được gọi là *trung bình Karcher* của các điểm $x_1, \dots, x_n$ nếu x là điểm cực tiểu của hàm

$$f(x) = \sum_{i=1}^n (d(x_i, x))^2 \quad (3.1)$$

trên M . Bài toán đi tìm trung bình Karcher được gọi là *bài toán K-mean*.

Ví dụ 3.1.2.

1. Ta tìm trung bình Karcher của hai điểm $A(a_1, \dots, a_n)$ và $B(b_1, \dots, b_n)$ trong không gian $\mathbb{R}^n$ với mêtric thông thường. Tức là cần tìm điểm cực tiểu của hàm số

$$f(x) = (x_1 - a_1)^2 + \dots + (x_n - a_n)^2 + (x_1 - b_1)^2 + \dots + (x_n - b_n)^2. \quad (3.2)$$

Áp dụng BĐT $a^2 + b^2 \geq \frac{(a+b)^2}{2}$, ta có

$$f(x) = (x_1 - a_1)^2 + (b_1 - x_1)^2 + \dots + (x_n - a_n)^2 + (b_n - x_n)^2$$

$$\geq \frac{(b_1 - a_1)^2}{2} + \dots + \frac{(b_n - a_n)^2}{2}.$$

Dấu đẳng thức xảy ra khi và chỉ khi $x = \left(\frac{a_1 + b_1}{2}, \dots, \frac{a_n + b_n}{2} \right)$. Tức là hàm số (3.2) đạt cực tiểu tại trung điểm của đoạn thẳng nối a và b .

2. Bây giờ ta tìm trung bình Karcher cho m điểm $A_1(a_{11}, \dots, a_{n1}), \dots, A_m(a_{1m}, \dots, a_{nm})$ trên mặt cầu đơn vị S^{n-1} . Do các điểm đều nằm trên mặt cầu nên

$$\sum_{i=1}^n x_i^2 = \sum_{i=1}^n a_{ik}^2 = 1, \quad k = 1, \dots, m.$$

Ta có:

$$\begin{aligned} f(x) &= \sum_{k=1}^m \sum_{i=1}^n (x_i - a_{ik})^2 \\ &= 2m - 2 \sum_{i=1}^n \left(\sum_{k=1}^m a_{ik} \right) x_i \\ &\geq 2m - 2 \sqrt{\sum_{i=1}^n x_i^2} \sqrt{\sum_{i=1}^n \left(\sum_{k=1}^m a_{ik} \right)^2} \\ &= 2m - 2 \sqrt{\sum_{i=1}^n \left(\sum_{k=1}^m a_{ik} \right)^2}. \end{aligned}$$

Dấu đẳng thức xảy ra khi và chỉ khi

$$\frac{x_1}{\sum_{k=1}^m a_{k1}} = \dots = \frac{x_n}{\sum_{k=1}^m a_{kn}}.$$

Đây là một tia đi qua gốc tọa độ và điểm trung bình số học $\bar{A} = \frac{1}{m}(A_1 + \dots + A_m)$.

Giao của tia này với S^{m-1} chính là điểm $x^* = \frac{\bar{A}}{\|\bar{A}\|}$ và đó cũng chính là cực tiểu của hàm f hay là nghiệm của bài toán K-mean (3.1) trên mặt cầu S^{m-1} với khoảng cách Euclide thông thường.

3.1.2 Thực hành với MATLAB

Trong mục này, chúng tôi lập trình một chương trình trong MATLAB để giải bài toán tìm trung bình Karcher của m điểm trên mặt cầu đơn vị S^2 trong $\mathbb{R}^3$.

Giả sử tọa độ của m điểm đó được cho dưới dạng các cột của ma trận

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ a_{31} & a_{32} & \dots & a_{3m} \end{pmatrix}.$$

Hàm mục tiêu khi đó là

$$\begin{aligned} f : S^2 &\rightarrow \mathbb{R} \\ x &\mapsto f(x) = \sum_{i=1}^m \|x - A[i]\|^2, \end{aligned}$$

với $A[i]$ là cột thứ i của A . Xét hàm số

$$\begin{aligned} \bar{f} : \mathbb{R}^3 &\rightarrow \mathbb{R} \\ x &\mapsto \bar{f}(x) = \sum_{i=1}^m \|x - A[i]\|^2, \end{aligned}$$

để thấy f là hạn chế của $\bar{f}$ trên S^2 . Ta có

$$\text{grad} \bar{f}(x) = 2mx - 2 \begin{pmatrix} a_{11} + a_{12} + \dots + a_{1m} \\ a_{21} + a_{22} + \dots + a_{2m} \\ a_{31} + a_{32} + \dots + a_{3m} \end{pmatrix} = 2mx - 2B,$$

ở đây B để là ma trận cỡ 3×1 mà $B_i = \sum_{k=1}^m a_{ik}$, $i = 1, 2, 3$.

Phép chiếu vuông góc lên không gian tiếp xúc $T_x S^2$ là $P_x \xi = (I - xx^\top) \xi$. Do đó

$$\text{grad} f(x) = P_x \text{grad} \bar{f}(x) = 2(I - xx^\top)(mx - B).$$

Để tính $\text{Hess} f(x)$, ta xét

$$\begin{aligned} D \text{grad} f(x)[\eta] &= 2 \lim_{t \rightarrow 0} \frac{(I - (x + t\eta)(x + t\eta)^\top)(m(x + t\eta) - B) - (I - xx^\top)(mx - B)}{t} \\ &= 2 \left(m(I - xx^\top)\eta - mx\eta^\top x + x\eta^\top B - m\eta x^\top x + \eta x^\top B \right). \end{aligned}$$

Từ đó, ta có

$$\text{Hess} f(x) = \nabla_\eta \text{grad} f(x)$$

$$\begin{aligned}
&= P_x(\text{Dgrad}f(x)[\eta]) \\
&= 2(I - xx^\top) \left(m(I - xx^\top)\eta - mx\eta^\top x + x\eta^\top B - m\eta x^\top x + \eta x^\top B \right).
\end{aligned}$$

Sử dụng $\eta \in T_x S^2$ và $(\eta x^\top + x\eta^\top)B = (xB^\top + x^\top B)\eta$, phương trình Newton (2.12) trở thành:

$$\begin{cases} 2(I - xx^\top) (m + xB^\top + x^\top B - mx^\top x) \eta = -2(I - xx^\top)(mx - B) \\ x^\top \eta = 0 \end{cases} \quad (3.3)$$

Chương trình dưới đây viết cho

$$A = \begin{pmatrix} \frac{1}{3} & 0 & 0 & \frac{1}{\sqrt{3}} \\ \frac{2}{3} & 1 & 0 & \frac{1}{\sqrt{3}} \\ \frac{2}{3} & 0 & 1 & \frac{1}{\sqrt{3}} \end{pmatrix}.$$

```

clear all
close all
disp('Tim trung binh Karcher tren mat cau bang phuong phap
    Newton:')
disp('Danh sach cac diem')
A = [1/3,0,0,1/sqrt(3);2/3,1,0,1/sqrt(3);2/3,0,1,1/sqrt(3)]
disp('Diem khoi tao')
x=[cos(pi/4);sin(pi/4);0]
ta = [1e-5;1e-6];
kmax = 20;
[k,valu,mi,err, nrmgrad,deri, tol] = ...
    Newton_Karcher(A,x,ta,kmax);
figure(1)
plot(1:k,err,'-b*',1:k,nrmgrad,':rs','LineWidth',2)
legend('sai so','chuan gradient')
figure(2)
plot(1:k,valu,'-k*', 'LineWidth',2)
legend('gia tri ham muc tieu')

```

Trong đó thủ tục Newton_Karcher(A,x,ta,kmax) là

```
function [k,valu,mi,err,nrmgrad,deri, tol] =
    Newton_Karcher(A,x,ta,kmax)
m= size(A,2);
mi=zeros(3,kmax);
err=zeros(kmax,1);
nrmgrad = zeros(kmax,1);
deri= zeros(kmax,1); % ghi lai gia tri gradient cua dao ham
valu = zeros(kmax,1); % gia tri ham muc tieu
%% Bat dau thuat toan Newton
r0 = norm(deriv(x));
deri(1) = r0;
err(1) = norm(x);
mi(:,1) = x;
nrmgrad(1) = norm(deriv(x));
valu(1) = ssd(x);
k=2;
tol = ta(1)*r0+ta(2);
while norm(deriv(x)) > tol && k < kmax
    x0 = x;
    s = -[ deriv2(x); x'] \ [deriv(x); 0] ;
    x = (1/norm(x+s))*(x+s) % Anh xa rut
    nrmgrad(k) = norm(deriv(x));
    err(k) = norm(x0-x);
    mi(:,k) = x;
    valu(k) = ssd(x);
    k = k+1;
end
if k == kmax
    disp('chuong trinh co the da that bai')
else
    valu(k) = ssd(x);
    mi = mi(:,1:k);
    err = err(1:k)
    valu = valu(1:k)
```

```

    nrmgrad = nrmgrad(1:k);
    deri = deri(1:k);
end
    disp('Trung binh Karcher can tim la')
x
    disp('Gia tri ham muc tieu')
ssd(x)

function y1 = ssd(x) % ham Binh phuong khoang cach
z = x;
for i=1:(m-1)
    z = [z,x];
end
    y1 = trace((z - A)*(z - A)');
end
function y2 = deriv(x)% gradient cua ham
y2 = 2*(eye(3) - x*x')*(m*x - sum(A,2));
end
function y3 = deriv2(x)% Hessian cua ham
y3 =2*(eye(3) - x*x')*(m*eye(3)+ x*(sum(A,2))' +
    x'*sum(A,2)*eye(3) - m*x'*x *eye(3));
end
end

```

Ta thu được kết quả như sau

Tim trung binh Karcher tren mat cau bang phuong phap Newton:

Danh sach cac diem

A =

0.3333	0	0	0.5774
0.6667	1.0000	0	0.5774
0.6667	0	1.0000	0.5774

Diem khoi tao

x =

```
0.7071
```

```
0.7071
```

```
0
```

```
Trung bình Karcher cần tìm là
```

```
x =
```

```
0.2758
```

```
0.6797
```

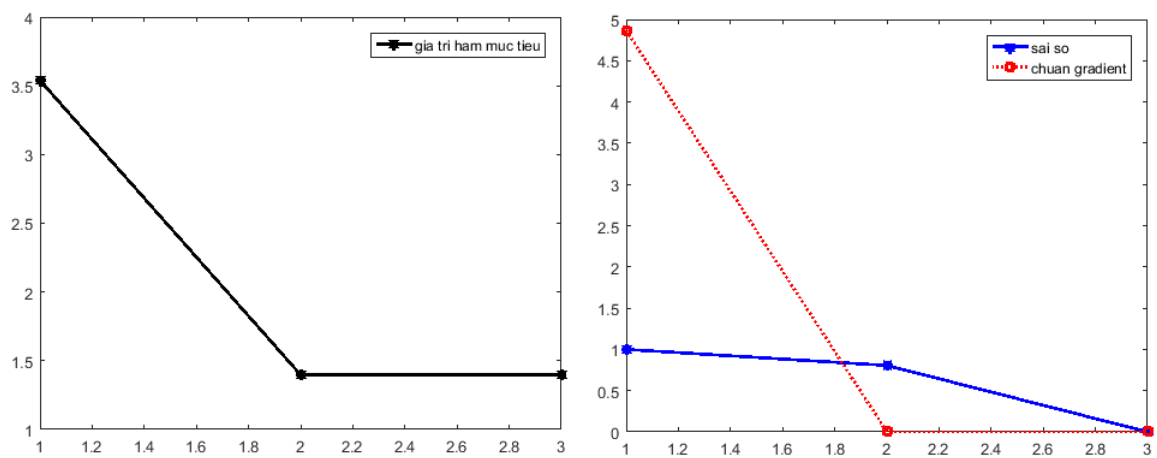
```
0.6797
```

```
Gia trị hàm mục tiêu
```

```
ans =
```

```
1.3968
```

Các biểu đồ



Hình 3.1: Biểu đồ giá trị hàm mục tiêu, sai số và chuẩn gradient với Bài toán K-mean.

3.2 Bài toán điểm trung chuyển hàng không

3.2.1 Giới thiệu

Năm 2015, dư luận trong nước rất quan tâm đến Dự án sân bay Long Thành. Trong Tờ trình [3] của Chính phủ và Nghị quyết [4] của Quốc hội đều nhắc đến việc Long Thành sẽ trở thành một cảng hàng không trung chuyển trong khu vực. Dễ dàng hình dung một trong những điều kiện thuận lợi cho một cảng hàng không trung chuyển là tổng khoảng cách từ sân bay đó đến các sân bay quốc tế trong khu vực là nhỏ nhất. Giả sử ta xây dựng một hệ trục tọa độ Descartes có tâm nằm ở tâm Trái đất. Khi đó, vị trí mỗi sân bay trong khu vực được đặc trưng bởi một bộ ba tọa độ thông thường trong không gian. Ta có thể thực hiện một phép co theo bán kính Trái đất để đưa bài toán đang xét về trên mặt cầu đơn vị. Việc kiểm tra vị trí sân bay Long Thành có đúng là một điểm trung chuyển tốt hay không hoàn toàn có thể được giải quyết dưới góc độ Toán học bằng cách xét bài toán tối ưu sau:

Cho n điểm $A_1, A_2, \dots, A_n$ trên cầu đơn vị S^2 , tìm vị trí điểm X sao cho tổng khoảng cách trắc địa từ X đến các điểm A_i là nhỏ nhất.

3.2.2 Bài toán điểm trung chuyển hàng không

Giả sử a, b là hai điểm trên mặt cầu S^2 trong $\mathbb{R}^3$. Ta tính khoảng cách trắc địa $d_g(a, b)$ từ a tới b . Áp dụng (1.9) ta có phương trình cung trắc địa đi qua a, b là

$$x(t) = a \cos(\|\xi\|t) + \frac{\xi}{\|\xi\|} \sin(\|\xi\|t),$$

với $\xi = P_a b = b - (a^\top b)a$, $\|\xi\| = \sqrt{1 - (a^\top b)^2}$. Dễ thấy

$$x(0) = a, x\left(\frac{\arccos(a^\top b)}{\sqrt{1 - (a^\top b)^2}}\right) = b.$$

Mặt khác

$$\dot{x}(t) = -a\|\xi\| \sin(\|\xi\|t) + \xi \cos(\|\xi\|t).$$

Áp dụng công thức (1.3), với $\theta = \frac{\arccos(a^\top b)}{\sqrt{1 - (a^\top b)^2}}$, ta có

$$\begin{aligned}
 d_g(a, b) &= \int_0^\theta \sqrt{\langle \dot{x}(t), \dot{x}(t) \rangle} dt \\
 &= \int_0^\theta \sqrt{\|\xi\|^2 - a^\top \xi \sin(2\|\xi\|t)} dt \\
 &= \int_0^\theta \|\xi\| dt && (\text{vì } \xi \perp a) \\
 &= \theta \|\xi\| \\
 &= \arccos(a^\top b).
 \end{aligned}$$

Vậy

$$d_g(a, b) = \arccos(a^\top b). \quad (3.4)$$

Giả sử có n điểm $A_i, i = 1, \dots, n$ trên mặt cầu đơn vị S^2 trong $\mathbb{R}^3$. Ta cần cực tiểu hóa hàm mục tiêu

$$\begin{aligned}
 f : S^2 &\rightarrow \mathbb{R} \\
 x &\mapsto \sum_{i=1}^n \arccos(x^\top A_i).
 \end{aligned}$$

Giả sử tọa độ của điểm A_i là cột thứ i của ma trận

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ a_{31} & a_{32} & \dots & a_{3n} \end{pmatrix}.$$

Trong mục này, với hàm số thực $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ và $x = (x_1, x_2, \dots, x_m)$, ta kí hiệu

$$\begin{aligned}
 \vec{\varphi} : \mathbb{R}^m &\rightarrow \mathbb{R}^m \\
 x &\mapsto \vec{\varphi}(x) := \begin{pmatrix} \varphi(x_1) \\ \varphi(x_2) \\ \dots \\ \varphi(x_m) \end{pmatrix}.
 \end{aligned}$$

Ta có

$$\text{grad} f(x) = -(I - xx^\top) A \vec{g}(x^\top A),$$

với $g(t) = \frac{1}{\sqrt{1-t^2}}, t \in (-1, 1)$. Do đó, theo (1.1)

$$\begin{aligned} \text{Dgrad}f(x)[\eta] &= \lim_{t \rightarrow 0} \frac{\text{grad}f(x+t\eta) - \text{grad}f(x)}{t} \\ &= (x\eta^\top + \eta x^\top) A \vec{g}(x^\top A) + (xx^\top - I)A \lim_{t \rightarrow 0} \frac{\vec{g}((x+t\eta)^\top A) - \vec{g}(x^\top A)}{t}. \end{aligned}$$

Ta lại có

$$\lim_{t \rightarrow 0} \frac{\vec{g}((x+t\eta)^\top A) - \vec{g}(x^\top A)}{t} = -(A^\top \odot H),$$

trong đó

$$H = \begin{pmatrix} \vec{h}(x^\top A), & \vec{h}(x^\top A), & \vec{h}(x^\top A) \end{pmatrix}, \quad h(t) = \frac{t}{\sqrt{(1-t^2)^3}}, \quad t \in (-1, 1)$$

và $\odot$ là phép nhân Hadamard hai ma trận cùng cỡ, tức là nhân phần tử với phần tử của hai ma trận cùng cỡ, với P, Q là các ma trận cỡ $m \times n$,

$$P \odot Q = \begin{pmatrix} p_{11}q_{11} & \dots & p_{1n}q_{1n} \\ \dots & \dots & \dots \\ p_{m1}q_{m1} & \dots & p_{mn}q_{mn} \end{pmatrix}.$$

Kết hợp với tính chất $(\eta x^\top + x\eta^\top)B = (xB^\top + x^\top B)\eta$, ta thu được

$$\text{Hess}f(x) = (I - xx^\top) \left(x \left(\vec{g}(x^\top A) \right)^\top A^\top + x^\top A \vec{g}(x^\top A) - (xx^\top - I)A(A^\top \odot H) \right) \eta.$$

3.2.3 Thực hành với MATLAB

Dưới đây, ta sẽ lập trình một chương trình tính toán vị trí điểm trung chuyển hàng không của 10 sân bay nhộn nhịp nhất trong khu vực Châu Á năm 2017 theo thống kê của *Airports Council International*.

Để đổi tọa độ địa lý sang tọa độ Descartes, ta sử dụng phép biến đổi

$$x = \cos(la) \cos(lo), \quad y = \cos(la) \sin(lo), \quad z = \sin(la),$$

trong đó (la, lo) là tọa độ địa lý.

STT	Tên sân bay	Tọa độ địa lý
1	Sân bay Quốc tế Bắc Kinh	(40.0725, 116.5975)
2	Sân bay Haneda	(35.553333, 139.781111)
3	Sân bay Quốc tế Hồng Kông	(22.308889, 113.914444)
4	Sân bay Quốc tế Phố Đông - Thượng Hải	(31.143333, 121.805278)
5	Sân bay Quốc tế Bạch Vân - Quảng Châu	(23.3925, 113.298889)
6	Sân bay Quốc tế Soekarno - Hatta	(-6.125556, 106.655833)
7	Sân bay Quốc tế Changi - Singapore	(1.359167, 103.989444)
8	Sân bay Quốc tế Incheon	(37.463333, 126.44)
9	Sân bay Quốc tế Suvarnabhumi	(13.6925, 100.75)
10	Sân bay Quốc tế Kuala Lumpur	(2.743333, 101.698056)

Bảng 3.1: Danh sách các sân bay nhộn nhịp nhất Châu Á năm 2017 (Theo *Airports Council International*).

Chương trình trong MATLAB dưới đây lấy điểm khởi tạo là (11, 106), trên địa phận Cambodia.

```

clear all
close all
disp('Bai toan Diem trung chuyen hang khong:')
disp('Danh sach cac san bay')
C = [ 40.0725, 116.5975 ; 35.553333, 139.781111; 22.308889,
      113.914444; 31.143333, 121.805278; 23.3925, 113.298889;
      -6.125556, 106.655833;
      1.359167, 103.989444; 37.463333, 126.44; 13.6925, 100.75;
      2.743333, 101.698056 ]' %Danh sach san bay theo toa do
      Dia ly
B = (pi/180)*C; % Doi don vi radian
m = size(B,2);
A = [cos(B(1,1))*cos(B(2,1));cos(B(1,1))*sin(B(2,1));
      sin(B(1,1)) ];
for i=2: m

```

```

A = [A, [cos(B(1,i))*cos(B(2,i));cos(B(1,i))*sin(B(2,i));
        sin(B(1,i))]];
end
disp('Danh sach san bay theo toa do Descartes')
A
disp('Diem khoi tao')
D=(pi/180)*[11, 106]' % Cambodia;
x = [cos(D(1,1))*cos(D(2,1));cos(D(1,1))*sin(D(2,1));
      sin(D(1,1)) ]
ta = [1e-2;1e-3];
kmax = 65 ;
[k,valu,mi,err, nrmgrad,deri, tol] = ...
    Newton_Airport(A,x,ta,kmax);
figure(1)
plot(1:k,err,'-b*',1:k,nrmgrad,':rs','LineWidth',2)
legend('sai so','chuan gradient')
figure(2)
plot(1:k,valu,'-k*','LineWidth',2)
legend('gia tri ham muc tieu')

```

Trong đó thủ tục Newton Airport(A,x,ta,kmax) là

```

function [k,valu,mi,err,nrmgrad,deri, tol] =
    Newton_Airport(A,x,ta,kmax)

fun = @(t) 1/sqrt(1-t^2);
fun2 = @(t) t/sqrt((1-t^2)^3);
mi=zeros(3,kmax);
err=zeros(kmax,1);
nrmgrad = zeros(kmax,1);
deri= zeros(kmax,1); % ghi lai gia tri gradient cua dao ham
valu = zeros(kmax,1); % gia tri ham muc tieu
%% Bat dau thuat toan Newton

r0 = norm(deriv(x));
deri(1) = r0;

```



```

err(1) = norm(x);
mi(:,1) = x;
nrmgrad(1) = norm(deriv(x));
valu(1) = airport(x);
k=2;
tol = ta(1)*r0+ta(2);

while norm(deriv(x)) > tol && k < kmax
    x0=x;
    s = -[ deriv2(x); x' ] \ [deriv(x); 0];
    x = (1/norm(x+s))*(x+s); % Anh xa rut
    nrmgrad(k) = norm(deriv(x));
    err(k) = norm(x0-x);
    mi(:,k) = x;
    valu(k) = airport(x);
    k = k+1;
end
if k == kmax
    disp('chuong trinh co the da that bai')
else
    valu(k) = airport(x);
    mi = mi(:,1:k);
    err = err(1:k);
    valu = valu(1:k);
    nrmgrad = nrmgrad(1:k);
    deri = deri(1:k-1);
end
disp('Diem Trung chuyen can tim la')
x
disp('Gia tri ham muc tieu')
airport(x)

function y1 = airport(x) % ham tong do dai duong bay
    y1 = sum(acos(x'*A),2);

```

```

end
function y2 = deriv(x)% gradient cua ham
y2 = -(eye(3)-x*x')*A* arrayfun(fun, (x'*A)');
end
function y3 = deriv2(x)% Hessian cua ham
h = arrayfun(fun2, (x'*A)');
H = [h, h, h];
y3 = (eye(3)-x*x')*(x*arrayfun(fun, (x'*A)')'*A' + x'*A*
    arrayfun(fun, (x'*A)')*eye(3) - (x*x' - eye(3))*A*(A'*H));
end
end

```

Ta thu được kết quả

Bai toan Diem trung chuyen hang khong:

Danh sach cac san bay

C =

```

40.0725 35.5533 22.3089 31.1433 23.3925 -6.1256 1.3592
37.4633 13.6925 2.7433
116.5975 139.7811 113.9144 121.8053 113.2989 106.6558
103.9894 126.4400 100.7500 101.6981

```

Danh sach san bay theo toa do Descartes

A =

```

-0.3426 -0.6212 -0.3750 -0.4511 -0.3630 -0.2850 -0.2417
-0.4715 -0.1812 -0.2025
0.6842 0.5253 0.8457 0.7274 0.8430 0.9526 0.9701
0.6385 0.9545 0.9781
0.6438 0.5815 0.3796 0.5172 0.3970 -0.1067 0.0237
0.6083 0.2367 0.0479

```

Diem khoi tao

x =

```

-0.2706
0.9436
0.1908

```

Điểm Trung chuyển cần tìm là

$x =$

-0.3740

0.8455

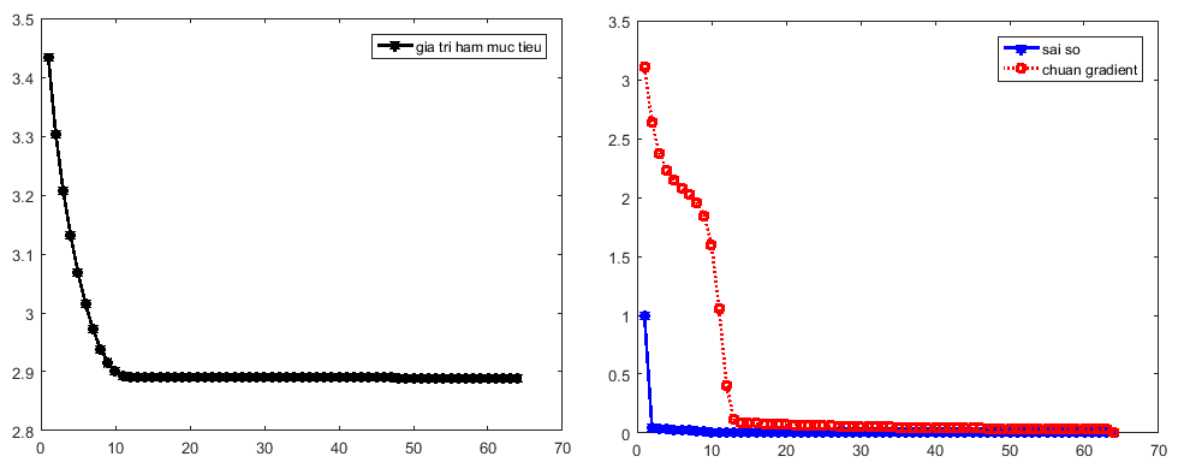
0.3811

Giá trị hàm mục tiêu

$ans =$

2.8898

Các biểu đồ



Hình 3.2: Biểu đồ giá trị hàm mục tiêu, sai số và chuẩn gradient với Bài toán điểm trung chuyển hàng không.

Đổi kết quả ra tọa độ địa lý, ta thu được điểm (22.4018, 113.8618), là một điểm gần sân bay Hồng Kông. Như vậy sân bay thuận lợi nhất cho việc trung chuyển giữa các sân bay trên là Sân bay Quốc tế Hồng Kông.

3.3 Bài toán giá trị riêng dưới góc độ tối ưu

Ta nhắc lại về giá trị riêng của một ma trận và các khái niệm liên quan. Giả sử $\mathbb{K}$ là trường số thực hoặc phức, A là ma trận cỡ $n \times n$ với các phần tử thuộc $\mathbb{K}$. Một vectơ khác không $v \in \mathbb{C}^n$ được gọi là *vectơ riêng* của A nếu tồn tại $\lambda \in \mathbb{C}$ thỏa mãn

$$Av = \lambda v.$$

Số λ khi đó được gọi là *giá trị riêng* của A . Cặp (λ, v) được gọi là *cặp riêng* của A . Tập hợp tất cả các giá trị riêng của A được gọi là *phổ* của A . Các giá trị riêng của A cũng là các nghiệm của *đa thức đặc trưng* của A ,

$$P_A(z) \equiv \det(A - zI).$$

Nếu T là một ma trận khả nghịch và (λ, v) là cặp riêng của A thì (λ, Tv) là một cặp riêng của TAT^{-1} .

Số λ là một giá trị riêng của ma trận A khi và chỉ khi $\dim \ker(A - \lambda I) > 0$. Khi đó, $\ker(A - \lambda I)$ được gọi là *không gian con riêng* của A ứng với λ .

Từ đây trở đi, ta chỉ xét các ma trận thực. Khi A là ma trận đối xứng cỡ $n \times n$ thì các giá trị riêng $\lambda_1 \leq \dots \leq \lambda_n$ của A là số thực và các vectơ riêng $v_1, v_2, \dots, v_n$ của nó có thể được chọn sao cho lập thành một hệ trực chuẩn, tức là

$$v_i^\top v_j = \begin{cases} 1 & \text{nếu } i = j \\ 0 & \text{nếu } i \neq j \end{cases}.$$

Nói cách khác, với mọi ma trận đối xứng A , tồn tại một ma trận trực giao V (với các cột là các vectơ riêng của A) và một ma trận đường chéo D sao cho $A = VDV^\top$. Giá trị riêng λ_1 được gọi là *giá trị riêng cực tả* của A và cặp (λ_1, v_1) được gọi là *cặp cực tả*. Một không gian con bất biến ứng với các giá trị riêng $\lambda_1, \dots, \lambda_p$ được gọi là *không gian con bất biến cực tả p -chiều*. Tương tự, *không gian con bất biến cực hữu p -chiều* là không gian con bất biến ứng với các giá trị riêng $\lambda_{n-p+1}, \dots, \lambda_n$. Không gian con bất

biến cực tả và không gian con bất biến cực hữu được gọi chung là *không gian con cực biên*.

3.3.1 Bài toán giá trị riêng dưới góc độ tối ưu

Ta nhắc lại kí hiệu $\text{tr}(A)$ để chỉ tổng các phần tử trên đường chéo của A .

Định nghĩa 3.3.1. Giả sử A là ma trận đối xứng cỡ $n \times n$, Y là ma trận hạng đủ cỡ $n \times p$. Khi đó hàm số

$$f(Y) = \text{tr} \left(Y^\top A Y (Y^\top Y)^{-1} \right) \quad (3.5)$$

được gọi là *thương Rayleigh tổng quát* của A .

Mệnh đề 3.3.2. Giả sử A là một ma trận đối xứng cỡ $n \times n$, $\lambda_1 \leq \dots \leq \lambda_n$ là các giá trị riêng của A . Khi đó các mệnh đề dưới đây là tương đương:

- (i) $\text{span}(Y^*)$ là một không gian con bất biến cực tả của A ;
- (ii) Y^* là một cực tiểu toàn cục của (3.5) trên tập tất cả ma trận hạng đủ cỡ $n \times p$;
- (iii) $f(Y^*) = \sum_{i=1}^p \lambda_i$.

Chứng minh. Để đơn giản, ta giả sử rằng $\lambda_p < \lambda_{p+1}$. Giả sử V là một ma trận cỡ $n \times n$ sao cho $V^\top V = I_n$ và $V^\top A V = \text{diag}(\lambda_1, \dots, \lambda_n)$. Dễ thấy V luôn tồn tại. Giả sử $Y \in \mathbb{R}^{n \times p}$ mà $Y = VM$. Vì $Y^\top Y = I_p$ nên $M^\top M = I_p$. Khi đó

$$\begin{aligned} \text{tr} (Y^\top A Y) &= \text{tr} (M^\top \text{diag}(\lambda, \dots, \lambda_n) M) \\ &= \sum_{i=1}^n \lambda_i \sum_{j=1}^p m_{ij}^2 \\ &= \sum_{j=1}^p \left(\lambda_p + \sum_{i=1}^p (\lambda_i - \lambda_p) m_{ij}^2 + \sum_{i=p+1}^n (\lambda_i - \lambda_p) m_{ij}^2 \right) \\ &= \sum_{i=1}^p \lambda_i + \sum_{i=1}^p (\lambda_p - \lambda_i) \left(1 - \sum_{i=1}^p m_{ij}^2 \right) + \sum_{j=1}^p \sum_{i=p+1}^n (\lambda_i - \lambda_p) m_{ij}^2. \end{aligned}$$

Vì số hạng thứ hai và cuối cùng không âm nên

$$\text{tr}(Y^\top AY) \geq \sum_{i=1}^p \lambda_i.$$

Dấu đẳng thức xảy ra nếu và chỉ nếu số hạng thứ hai và số hạng cuối cùng bằng 0. Điều này xảy ra khi và chỉ khi mảng $p \times p$ bên trên của M là trực giao và mảng $(n-p) \times p$ bên dưới của M chứa toàn phần tử 0. Tức là $Y = VM$ sinh một không gian con bất biến cực tả p -chiều của A . $\square$

Trong trường hợp $p = 1$, giả sử giá trị riêng cực tả λ_1 của A có bội 1. Mệnh đề 3.3.2 cho ta cực tiểu toàn cục của hàm mục tiêu

$$\begin{aligned} f : \mathbb{R}_*^n &\rightarrow \mathbb{R} \\ y &\mapsto f(y) = \frac{y^\top Ay}{y^\top y} \end{aligned} \quad (3.6)$$

là điểm $v_1 r$, $r \in \mathbb{R}_*$, với v_1 là vectơ riêng ứng với λ_1 . Tối ưu hóa thương Rayleigh (3.6) có thể được coi là một bài toán tối ưu trên đa tạp vì $\mathbb{R}_*^n$ có thể coi là một đa tạp với cấu trúc tự nhiên. Tuy nhiên, khía cạnh đa tạp ít được quan tâm, vì đó đơn giản chỉ là một không gian tuyến tính cổ điển bỏ đi phần tử 0.

Một vấp đề nữa là nghiệm tối ưu của bài toán tuy không bị cô lập nhưng đạt tới continuum $v_1 \mathbb{R}_*$. Do đó một số kết quả hội tụ quan trọng của các phương pháp tối ưu không thể áp dụng được cũng như một vài thuật toán quan trọng có thể sai.

3.3.2 Thuật toán thương Rayleigh trên mặt cầu

Để khắc phục tình trạng trên, người ta xét bài toán trên mặt cầu. Nghĩa là ta ràng buộc điều kiện có chuẩn 1 cho các vectơ riêng. Về khía cạnh tính toán, điều này được khuyến khích vì nó loại trừ được việc chia cho một đại lượng có thể có giá trị rất nhỏ. Theo đó, ta áp dụng Thuật toán 3 cho thương Rayleigh trên mặt cầu

$$\begin{aligned} f : S^{n-1} &\rightarrow \mathbb{R} \\ x &\mapsto x^\top Ax, \end{aligned} \quad (3.7)$$

ở đây ma trận A là một ma trận đối xứng. Xét hàm mục tiêu

$$\begin{aligned}\bar{f} : \mathbb{R}^n &\rightarrow \mathbb{R} \\ x &\mapsto x^\top Ax.\end{aligned}$$

Dễ thấy hạn chế của $\bar{f}$ trên mặt cầu S^{n-1} chính là (3.7).

Ta coi S^{n-1} là đa tạp Riemann con của không gian Euclide $\mathbb{R}^n$ được trang bị mêtric Riemann chuẩn tắc

$$\bar{g}(\xi, \zeta) = \xi^\top \zeta.$$

Với $x \in S^{n-1}$, ta có

$$D\bar{f}(x)[\zeta] = \zeta^\top Ax + x^\top A\zeta = 2\zeta^\top Ax,$$

với mọi $\zeta \in T_x \mathbb{R}^n \simeq \mathbb{R}^n$. Từ đó ta có thể định nghĩa

$$\text{grad} \bar{f}(x) = 2Ax.$$

Không gian tiếp xúc của S^{n-1} là $T_x S^{n-1} = \{\xi \in \mathbb{R}^n : x^\top \xi = 0\}$. Phép chiếu vuông góc lên không gian tiếp xúc tại x là

$$P_x \xi = \xi - xx^\top \xi.$$

Chú ý rằng $\text{grad} f(x) = P_x \text{grad} \bar{f}(x)$, ta được

$$\text{grad} f(x) = 2P_x(Ax) = 2(Ax - xx^\top Ax). \quad (3.8)$$

Các công thức trên được tổng hợp trong Bảng 3.2.

Mệnh đề 3.3.3. *Giả sử A là một ma trận đối xứng cỡ $n \times n$ và $x \in \mathbb{R}^n$ (với $\|x\| = 1$) là một vectơ riêng của A khi và chỉ khi nó là một điểm tới hạn của thương Rayleigh (3.7).*

Chứng minh. Giả sử x là một điểm tới hạn của (3.7), tức là $\text{grad} f(x) = 0$ với $x \in S^{n-1}$.

Từ (3.8) suy ra x thỏa mãn

$$Ax = (x^\top Ax) x = \lambda x.$$

	S^{n-1}	$\mathbb{R}^n$
hàm mục tiêu	$f(x) = x^\top Ax$	$\bar{f}(x) = x^\top Ax$
mêtric	mêtric cảm sinh	$\bar{g}(\xi, \zeta) = \xi^\top \zeta$
không gian tiếp xúc	$\xi \in \mathbb{R}^n : x^\top \xi = 0$	$\mathbb{R}^n$
phép chiếu vuông góc lên không gian tiếp xúc	$P_x \xi = (I - xx^\top) \xi$	phép đồng nhất
gradient	$\text{grad} f(x) = P_x \text{grad} \bar{f}(x)$	$\text{grad} \bar{f}(x) = 2Ax$
ánh xạ rút	$R_x(\xi) = \frac{x + \xi}{\ x + \xi\ }$	$R_x(\xi) = x + \xi$

Bảng 3.2: Thương Rayleigh trên mặt cầu.

Ngược lại, nếu x là một vector riêng (có chuẩn bằng 1) của A thì tồn tại một vô hướng λ sao cho

$$Ax = \lambda x.$$

Nhân bên trái cả hai vế của đẳng thức trên với $x^\top$, ta có $\lambda = x^\top$ và do đó $Ax = (x^\top Ax) x$. Vậy $\text{grad} f(x) = 0$. $\square$

Mệnh đề 3.3.4. *Giả sử A là ma trận đối xứng cỡ $n \times n$ với các giá trị riêng $\lambda_1 \leq \dots \leq \lambda_n$ và hệ vectơ riêng trực chuẩn tương ứng $v_1, \dots, v_n$. Khi đó*

- (i) $\pm v_1$ là cực tiểu địa phương và toàn cục của thương Rayleigh (3.7); nếu giá trị riêng λ_1 là bội 1 thì chúng là cực tiểu duy nhất.
- (ii) $\pm v_n$ là cực đại địa phương và toàn cục của thương Rayleigh (3.7); nếu giá trị riêng λ_n là bội 1 thì chúng là cực đại duy nhất.
- (iii) $\pm v_q$ tương ứng với giá trị riêng λ_q ($\lambda_1 < \lambda_q < \lambda_n$) là các điểm yên ngựa của (3.7).

Chứng minh.

- (i) được suy ra từ Mệnh đề 3.3.2.

(ii) được suy ra từ mệnh đề tương tự Mệnh đề 3.3.2 bằng cách thay A bởi $-A$, "cực tiểu" bởi "cực đại", "cực tả" bởi "cực hữu".

(iii) Xét cung tham số $\gamma : t \mapsto \frac{v_q + tv_1}{\|v_1 + tv_1\|}$. Dễ thấy

$$\frac{d^2}{dt^2}(f(\gamma(t)))|_{t=0} = \lambda_1 - \lambda_q < 0.$$

Tương tự, với cung tham số $\gamma : t \mapsto \frac{v_q + tv_n}{\|v_1 + tv_n\|}$, ta có

$$\frac{d^2}{dt^2}(f(\gamma(t)))|_{t=0} = \lambda_n - \lambda_q > 0.$$

Do đó, v_q là một điểm yên ngựa của thương Rayleigh (3.7).

□

Mệnh đề 3.3.3 và Định lý 2.2.10 suy ra rằng mọi phương pháp dựa trên Thuật toán 3 đều xây dựng dãy hội tụ về tập các vectơ riêng của A .

Sau đây ta áp dụng Thuật toán 3 vào bài toán giá trị riêng trên mặt cầu với hàm mục tiêu (3.7). Khi đó, ta có

$$\eta_k := -\text{grad} f(x_k) = 2(Ax_k - x_k x_k^\top Ax_k).$$

Dễ dàng chọn hướng tìm kiếm là dãy liên kết gradient và ánh xạ rút là

$$R_x(\xi) := \frac{x + \xi}{\|x + \xi\|}, \quad (3.9)$$

ở đây $\|\cdot\|$ là chuẩn Euclide trong $\mathbb{R}^n$, $\|y\| := \sqrt{y^\top y}$. Một lựa chọn khác là

$$R_x(\xi) := x \cos \|\xi\| + \frac{\xi}{\|\xi\|} \sin \|\xi\|, \quad (3.10)$$

với cung tham số $t \mapsto R_x(t\xi)$ là đường tròn lớn của mặt cầu. Ánh xạ rút thứ hai là ánh xạ mũ.

Vì S^{n-1} là một đa tạp Riemann của không gian Euclide $\mathbb{R}^n$, nên

$$\nabla_\eta \xi = P_x(D\xi(x)[\eta])$$

với mọi η thuộc không gian tiếp xúc $T_x S^{n-1}$ và mọi trường vectơ ξ trên S^{n-1} .

Bây giờ ta áp dụng Thuật toán 4 cho trường vectơ $\xi := \text{grad} f$, với f là thương Rayleigh (3.7). Với mọi $\eta \in T_{x_k} S^{n-1}$, ta có

$$\begin{aligned} \nabla_\eta \text{grad} f(x) &= 2P_x(\text{Dgrad} f(x)[\eta]) \\ &= 2P_x(A\eta - \eta x^\top A x) \\ &= 2(P_x A P_x \eta - \eta x^\top A x), \end{aligned}$$

ở đây $P_x x = 0$ và $P_x \eta = \eta$. Do đó, phương trình Newton (2.12) trở thành

$$\begin{cases} P_x A P_x \eta - \eta x^\top A x = -P_x A x, \\ x^\top \eta = 0, \end{cases}$$

ta có thuật toán sau

Algorithm 5 Phương pháp Newton cho thương Rayleigh trên mặt cầu S^{n-1}

Require: Ma trận đối xứng A .

Input: Giá trị ban đầu $x_0 \in \mathcal{M}$.

Output: Dãy $\{x_k\}$.

- 1: **for** $k = 0, 1, 2, \dots$ **do**
- 2: Giải hệ phương trình tuyến tính

$$\begin{cases} P_{x_k} A P_{x_k} \eta_k - \eta_k x_k^\top A x_k = -P_{x_k} A x_k, \\ x_k^\top \eta_k = 0, \end{cases} \quad (3.11)$$

với ẩn $\eta_k \in \mathbb{R}^n$.

- 3: Đặt $x_{k+1} := R_x(\eta_k)$.
 - 4: **end for**
-

3.3.3 Thực hành với MATLAB

```

clear all
close all
disp('Bai toan Thuong Rayleigh tren mat cau bang phuong phap
    Newton:')
disp('Ma tran doi xung')
A = [1,2,3,4;2,4,5,6;3,5,6,7;4,6,7,8]
disp('Diem khoi tao')
x = [0;1;0;0]
% x=[1/sqrt(3);1/sqrt(3);-1/sqrt(3)];
%x=[1/sqrt(3);1/sqrt(3);1/sqrt(3)]
ta = [1e-5;1e-6];
kmax = 20;
[k,valu,mi,err, nrmgrad,deri, tol] = ...
    Newton_GTR(A,x,ta,kmax);
disp('Vector rieng va gia tri rieng (de kiem tra)')
[V,d] = eig(A)
figure(1)
plot(1:k,err,'-b*',1:k,nrmgrad,':rs','LineWidth',2)
legend('sai so','chuan gradient')
figure(2)
plot(1:k,valu,'-k*','LineWidth',2)
legend('gia tri ham muc tieu')

```

Trong đó thủ tục `Newton_GTR(A,x,ta,kmax)` là

```

function [k,valu,mi,err,nrmgrad,deri, tol] =
    Newton_GTR(A,x,ta,kmax)

n = size(A,1);
mi=zeros(n,kmax);
err=zeros(kmax,1);
nrmgrad = zeros(kmax,1);
deri= zeros(kmax,1); % ghi lai gia tri gradient cua dao ham
valu = zeros(kmax,1); % gia tri ham muc tieu

```

```

%% Bat dau thuat toan Newton
r0 = norm(deriv(x));
deri(1) = r0;
err(1) = norm(x);
mi(:,1) = x;
nrmgrad(1) = norm(deriv(x));
valu(1) = Rq(x);
k=2;
tol = ta(1)*r0+ta(2);
while norm(deriv(x)) > tol && k < kmax
    x0=x;
    s = -[ deriv2(x); x' ] \ [deriv(x); 0] ;
    x = (1/norm(x+s))*(x+s); % Anh xa rut
    nrmgrad(k) = norm(deriv(x));
    err(k) = norm(x0-x);
    mi(:,k) = x;
    valu(k) = Rq(x);
    k = k+1;
end
if k == kmax
    disp('chuong trinh co the da that bai')
else
    valu(k) = Rq(x);
    mi = mi(:,1:k);
    err = err(1:k);
    valu = valu(1:k);
    nrmgrad = nrmgrad(1:k);
    deri = deri(1:k);
end
disp('Vector rieng')
x
disp('Gia tri ham muc tieu')
Rq(x)
function y1 = Rq(x) % Thuong Rayleigh
    y1 = x'*A*x;

```

```

end
function y2 = deriv(x)% gradient cua ham
y2 = (eye(n)-x*x')*A*x;
end
function y3 = deriv2(x)% Hessian cua ham
y3 =(eye(n) - x*x')*A*(eye(n) - x*x')-x'*A*x*eye(n);
end
end

```

Ta thu được kết quả

Bài toán Thương Rayleigh trên ma trận đối xứng
Newton:

Ma trận đối xứng

A =

1	2	3	4
2	4	5	6
3	5	6	7
4	6	7	8

Điểm khởi tạo

x =

0
1
0
0

Vector riêng

x =

-0.6317
0.6742
0.1644
-0.3454

Giá trị hàm mục tiêu

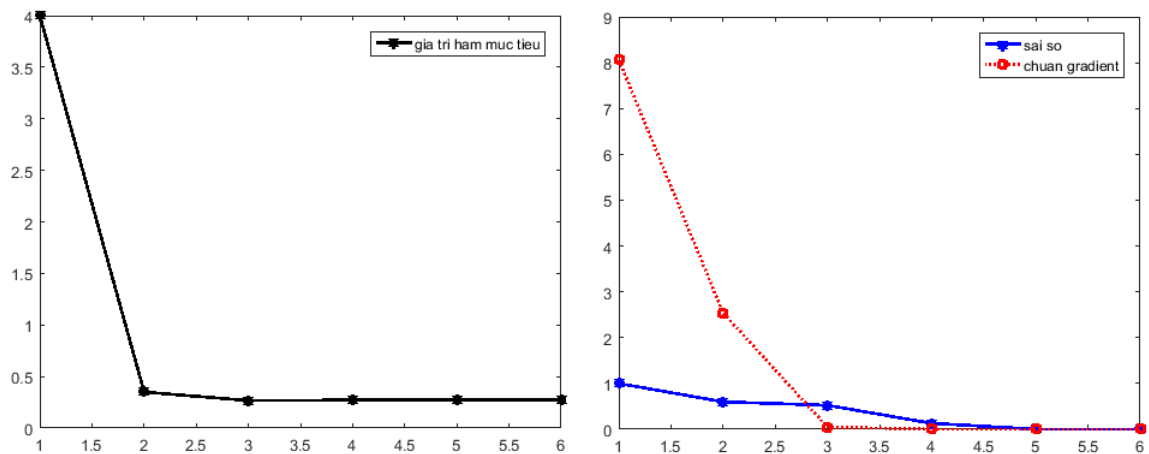
```
ans =
    0.2716
```

Vector riêng và giá trị riêng (để kiểm tra)

```
V =
    0.7256 -0.0000 -0.6317 0.2729
    0.4167 0.4082 0.6742 0.4529
   -0.0636 -0.8165 0.1644 0.5498
   -0.5439 0.4082 -0.3454 0.6467
```

```
d =
   -1.1132         0         0         0
         0    0.0000         0         0
         0         0    0.2716         0
         0         0         0   19.8415
```

Các biểu đồ



Hình 3.3: Biểu đồ giá trị hàm mục tiêu, sai số và chuẩn gradient với Bài toán thương Rayleigh trên mặt cầu.

Kết luận

Trong luận văn này, chúng tôi đã trình một số khái niệm, tính chất của đa tạp Riemann và thuật toán Tìm theo đường thẳng, phương pháp Newton trên đa tạp Riemann. Đặc biệt, chúng tôi chú trọng việc áp dụng thuật toán đã trình bày để giải quyết các bài toán K-mean, bài toán điểm trung chuyển hàng không, bài toán giá trị riêng và lập trình tính toán trên phần mềm MATLAB.

Tài liệu tham khảo

Tiếng Việt

- [1] Khu Quốc Anh, Nguyễn Doãn Tuấn (2005), *Lí thuyết liên thông và hình học Riemann*, NXB Đại học Sư Phạm.
- [2] Đoàn Quỳnh (2006), *Hình học vi phân*, NXB Đại học Sư Phạm.
- [3] Chính phủ nước CHXHCN Việt Nam, *Tờ trình đầu tư dự án xây dựng cảng hàng không quốc tế Long Thành*, 360/TTr-CP ngày 01 tháng 10 năm 2014.
- [4] Quốc hội nước CHXHCN Việt Nam, *Nghị quyết về chủ trương đầu tư dự án cảng hàng không quốc tế Long Thành*, 94/2015/QH13.

Tiếng Anh

- [5] Absil P.-A., Mahony R., Sepulchre R. (2008), *Optimization Algorithms on Matrix Manifolds*, Princeton University Press.
- [6] Kelly C. T. (1999), *Iterative Methods for Optimization*, SIAM, Philadelphia.