

UNIVERSIDAD NACIONAL DEL CALLAO

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

ESCUELA PROFESIONAL DE MATEMÁTICA



**MÉTODO DE MÁXIMO DESCENSO USANDO
RETRACCIONES EN VARIEDADES RIEMANNIANAS**

TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE
LICENCIADO EN MATEMÁTICA

EVER FRANKLIN CRUZADO QUISPE

Callao, Diciembre, 2014

PERÚ

HOJA DE REFERENCIA DEL JURADO Y APROBACIÓN

Método de Máximo Descenso Usando Retracciones en Variedades Riemannianas.

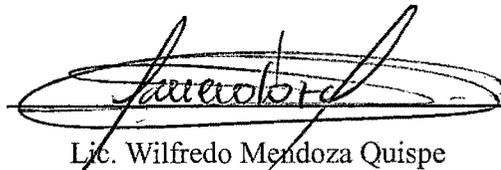
EVER FRANKLIN CRUZADO QUISPE

Tesis presentada a consideración del cuerpo docente de la Facultad de Ciencias Naturales y Matemática de la Universidad Nacional del Callao, como parte de los requisitos para obtener el título profesional de Licenciado en Matemática.

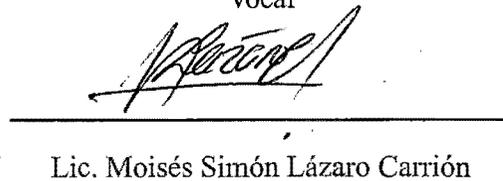
Aprobada por


Lic. Ezequiel Francisco Fajardo Campos

Presidente


Lic. Wilfredo Mendoza Quispe

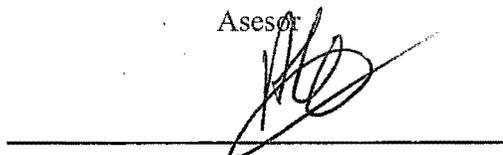
Vocal


Lic. Moisés Simón Lázaro Carrión

Secretario


Lic. Absalón Castillo Valdivieso

Asesor



Dr. Erik Alex Papa Quiroz

Co-Asesor

A mi madre:

*Por haberme brindado en todo momento
su apoyo y consejos para hacer de mi
una mejor persona.*

AGRADECIMIENTOS

Al concluir este trabajo, que presento como tesis para obtener el Título profesional de Matemática, en primer lugar doy gracias a Dios por haberme permitido llegar a este momento tan especial de mi vida y por darme la fortaleza para superar obstáculos y dificultades a lo largo de toda mi vida.

A mi Madre, por haberme formado con buenos sentimientos, hábitos y valores. Además quiero agradecerle por la confianza depositada en mí, por demostrarme siempre su cariño y apoyo incondicional, convirtiéndose de esta manera en el pilar más importante de mi vida.

A mi familia en general, que me han mostrado siempre una confianza infinita y me han dado su apoyo incondicional, les doy las gracias a todos ellos, a mis tíos, primos y abuelitos, que son para mí como mis padres, y demás familiares aunque algunos de ellos ya no están físicamente conmigo, estarán siempre presentes en mis recuerdos y en mi corazón, y en especial a mi hermana Rocío, que a pesar de su temprana partida, desde el cielo me protege y guía para que todo salga bien.

Al profesor Dr. Erik Alex Papa Quiroz, por haber sido un excelente orientador, por su visión crítica de muchos aspectos cotidianos de la vida, por sus enseñanzas y consejos que ayudan a formarte como persona y como profesional. También agradezco por su apoyo, motivación y su gran experiencia como investigador, lo cual fue muy importante para la realización de este trabajo.

A mi asesor Lic. Absalón Castillo Valdivieso, por acreditar en mi trabajo, también por su paciencia y tiempo que me ha brindado para la culminación con éxito de este trabajo.

A todos mis amigos, aunque ajenos a este trabajo han ayudado de manera directa o indirecta para su realización, en especial a todos mis amigos y compañeros que conocí en la FCNM-UNAC, por la amistad, los buenos momentos que pasamos, y su contribución y aliento mutuo para seguir adelante en esta hermosa carrera.

ÍNDICE

TABLAS DE CONTENIDO	4
RESUMEN	5
ABSTRACT	6
CAPÍTULO I: PLANTEAMIENTO DE LA INVESTIGACIÓN	7
1. Identificación del problema	7
2. Formulación del problema	8
3. Objetivos de la investigación	9
4. Justificación	9
5. Importancia	10
CAPÍTULO II: MARCO TEÓRICO	11
1. Preliminares	11
1.1 Simbología y notaciones	11
1.2 Nociones básicas de la teoría de variedades	11
1.3 Topología de una variedad	14
1.4 Funciones diferenciables	17
1.4.1 Inmersiones y submersiones	18
1.5 Subvariedades regulares	19
1.5.1 La variedad de Stiefel	20
1.6 Vectores tangentes y aplicaciones diferenciables	22
1.6.1 Vectores tangentes	23
1.6.2 Vectores tangentes a un espacio vectorial	27

1.6.3	Fibrado tangente	28
1.6.4	Campos vectoriales	28
1.6.5	Diferencial de una aplicación	29
1.6.6	Vectores tangentes a subvariedades regulares	30
1.7	Métrica riemanniana, distancia y gradientes	34
1.7.1	Subvariedades riemannianas	36
2.	Método de máximo descenso usando retracciones en variedades riemannianas	38
2.1	Retracciones	38
2.1.1	Retracciones sobre subvariedades regulares	40
2.2	Método del máximo descenso	43
2.3	Análisis de convergencia	45
2.3.1	Convergencia sobre variedades	45
2.3.2	Convergencia de los métodos de búsqueda lineal	46
2.4	Estabilidad de puntos fijos	48
2.5	Rapidez de convergencia	50
2.5.1	Orden de convergencia	50
2.6	Tasa de convergencia de los métodos de búsqueda lineal	52
3.	Aplicaciones del método de máximo descenso usando retracciones en variedades riemannianas	56
3.1	Minimización del cociente de Rayleigh sobre la esfera unitaria	56
3.1.1	Función objetivo y cálculo del gradiente	56
3.1.2	Puntos críticos del cociente de Rayleigh	58
3.1.3	Búsqueda lineal de Armijo	60
3.1.4	Implementación del Algoritmo 2 en MATLAB	60
3.2	Minimización de la función de Brockett sobre la variedad de Stiefel	65
3.2.1	Función objetivo y dirección de Búsqueda	65
3.2.2	Implementación del algoritmo que minimiza la función de Brockett sobre la variedad de Stiefel	67

CAPÍTULO III: VARIABLES E HIPÓTESIS	71
1. Variable de la investigación	71
2. Operacionalización de variables	71
3. Hipótesis general e hipótesis específica	72
CAPÍTULO IV: METODOLOGÍA	73
1. Tipo de investigación	73
2. Diseño de la investigación	73
3. Población y muestra	74
4. Técnicas e instrumentos de recolección de datos	74
5. Procedimiento de recolección de datos	74
6. Procesamiento estadístico y análisis de datos	74
CAPÍTULO V: RESULTADOS	75
CAPÍTULO VI: DISCUSIÓN DE RESULTADOS	76
CAPÍTULO VII: CONCLUSIONES	77
CAPÍTULO VIII: RECOMENDACIONES	78
BIBLIOGRAFÍA	79
ANEXOS	82

TABLAS DE CONTENIDO

Índice de figuras

Figura 2.1: Carta	12
Figura 2.2: Cartas y cambio de coordenadas.	13
Figura 2.3: Funciones diferenciables	18
Figura 2.4: Aplicación diferencial de F en x	30
Figura 2.5: Curvas y vectores tangentes sobre la esfera S^{n-1}	31
Figura 2.6: Espacio tangente sobre la esfera S^{n-1}	33
Figura 2.7: Retracción	39

Índice de tablas

Tabla 2.1: Cociente de Rayleigh sobre la esfera unitaria	57
Tabla 2.2: Resultados numéricos de minimizar el cociente de Rayleigh	63
Tabla 2.3: Función de Brockett sobre la variedad de Stiefel	66
Tabla 2.4: Resultados numéricos de minimizar la función de Brockett	70

RESUMEN

MÉTODO DE MÁXIMO DESCENSO USANDO RETRACCIONES EN VARIEDADES RIEMANNIANAS

Ever Franklin Cruzado Quispe

Diciembre - 2014

Asesor: Lic. Absalón Castillo Valdivieso

Título obtenido: Licenciado en Matemática

En el presente trabajo mostraremos un algoritmo de búsqueda lineal para el método de máximo descenso en variedades riemannianas usando retracciones. Específicamente estudiamos la búsqueda lineal de Armijo para llegar a un punto mínimo de una función objetivo dada. También estudiaremos la convergencia y velocidad de convergencia del método de máximo descenso para esta búsqueda lineal. Luego aplicamos el método estudiado para minimizar el cociente de Rayleigh sobre la esfera unitaria y para minimizar la función de Brockett sobre la variedad de Stiefel, de los cuales mostramos e implementamos su algoritmo en MATLAB, siendo este nuestro objetivo principal. Además presentaremos los resultados numéricos obtenidos en MATLAB.

Palabras Claves: Optimización en variedades, variedades riemannianas, retracciones, método de máximo descenso.

ABSTRACT

METHOD OF STEEPEST DESCENT USING RETRACTIONS IN RIEMANNIAN MANIFOLDS

Ever Franklin Cruzado Quispe

December - 2014

Adviser: Lic. Absalón Castillo Valdivieso

Degree obtained: Licentiate in Mathematics

In this work, we will develop a algorithm of line search for the method of steepest descent using retractions in riemannian manifolds. In specific we will study the Armijo line search for obtain a minimum point of a given cost function. Also we will study the convergence and speed of convergence of the method of steepest descent for this line search. Then we apply the studied method for minimize the Rayleigh quotient on the sphere and for minimize the brockett function on the Stiefel manifold, of wich we show and implement your algorithm en MATLAB, being this our main objective. Else we will present the numerical results obtained in MATLAB.

Palabras Claves: Optimization on manifolds, riemannian manifolds, retractions, steepest descent method.

CAPÍTULO I

PLANTEAMIENTO DE LA INVESTIGACIÓN

1. Identificación del problema

El modelo general que representa a un problema de optimización es el siguiente:

$$\begin{cases} \min f(x) \\ s.a. \\ x \in X \subseteq \mathbb{R}^n \end{cases} \quad (1.1)$$

donde, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es la función objetivo, X es el conjunto de restricciones del problema (1.1) y $x \in X$ es la variable que debe ser determinada. Si en el problema (1.1) no se tiene ninguna restricción para x , es decir $X = \mathbb{R}^n$, el modelo se llama “Problema de Optimización sin Restricciones”.

El modelo (1.1) se puede generalizar para otros espacios como los espacios de Banach, Hilbert y también para variedades riemannianas; si el modelo (1.1) lo generalizamos para variedades riemannianas obtendremos el siguiente modelo:

$$\begin{cases} \min f(x) \\ s.a. \\ x \in \mathcal{M} \end{cases} \quad (1.2)$$

donde, f es una función definida en una variedad riemanniana \mathcal{M} , es decir, $f : \mathcal{M} \rightarrow \mathbb{R}$.

Para resolver el modelo (1.1) existen muchos métodos iterativos uno de ellos es el método de máximo descenso el cual genera una sucesión de puntos x_k dado por:

$$x_{k+1} = x_k + \lambda_k d_k,$$

donde d_k es la dirección de descenso, es decir, una dirección que cumple $\nabla f(x_k)^T d_k < 0$, para el método de máximo descenso $d_k = -\nabla f(x_k)$, puesto que esta dirección negativa del gradiente me dá el mayor decrecimiento y λ_k es la longitud de paso, el cual se determinará por la regla de Armijo o por resolver un

subproblema de minimización llamado búsqueda lineal, es por eso que este tipo de métodos se denomina algoritmos de búsqueda lineal.

Si el método de máximo descenso se generaliza para resolver el problema (1.2), se denominará método de máximo descenso en variedades riemannianas, pero ahora la sucesión de puntos x_k que se genera para llegar al óptimo, esta dado por:

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

donde $\eta_k \in T_{x_k} \mathcal{M}$ ($T_{x_k} \mathcal{M}$ es el espacio tangente a la variedad \mathcal{M} en el punto x_k), t_k es un escalar para el cual se hará una búsqueda de Armijo para determinarlo y R es una aplicación llamada **retracción**, esta aplicación se define como una aplicación suave del fibrado tangente sobre la variedad, es decir, $R : T\mathcal{M} \rightarrow \mathcal{M}$; además y si denotamos por R_x la restricción de R a $T_x \mathcal{M}$ se debe cumplir que:

- $R_x(0_x) = x$, donde 0_x denota el elemento cero de $T_x \mathcal{M}$.
- Con la identificación canónica $T_{0_x} T_x \mathcal{M} \simeq T_x \mathcal{M}$, R_x satisface $DR_x(0_x) = \text{id}_{T_x \mathcal{M}}$, donde, $\text{id}_{T_x \mathcal{M}}$ denota la aplicación identidad sobre $T_x \mathcal{M}$.

Finalmente, una vez estudiado el método de máximo descenso en variedades riemannianas, surge la interrogante, ¿será posible implementar en MATLAB dicho método de manera análoga como en el espacio euclidiano?. En ese sentido el presente proyecto está destinado a estudiar la aplicación retracción para así poder aplicarlo al método de máximo descenso sobre variedades riemannianas, así como también estudiar la convergencia, velocidad de convergencia del método y finalmente se presentará la implementación del método para algunos problemas particulares; para lo cual usaremos algunas herramientas de la geometría riemanniana.

2. Formulación del problema

Se pretende resolver y analizar las siguientes interrogantes:

1. ¿Será posible aplicar el método de máximo descenso usando retracciones, para resolver problemas de optimización en variedades riemannianas?

2. ¿Será posible implementar el método de máximo descenso usando retracciones en variedades riemannianas usando MATLAB?

3. Objetivos de la investigación

Objetivo general

Estudiar las propiedades de convergencia del método de máximo descenso usando retracciones en variedades riemannianas, plantear su algoritmo e implementarlo en MATLAB.

Objetivos específicos

1. Estudiar la aplicación retracción y determinarlo para algunas variedades particulares.
2. Estudiar la extensión del método de máximo descenso usando retracciones en variedades riemannianas.
3. Minimizar el cociente de Rayleigh sobre la esfera unitaria.
4. Minimizar la función de Brocket sobre la variedad de Stiefel.

4. Justificación

Nuestra justificación para realizar este estudio radica en que los algoritmos de búsqueda lineal, utilizan una dirección de descenso, esto se convierte en algunos casos en un proceso un tanto tedioso al implementar el algoritmo. Es por eso que estamos interesados, no solo en estudiar el método de máximo descenso en variedades riemannianas, su convergencia y velocidad de convergencia sino que implementaremos su algoritmo en MATLAB de manera que sea un procedimiento numérico eficiente en la solución de problemas de optimización.

5. Importancia

En la actualidad la Optimización Matemática es de gran interés puesto que es aplicado a muchos problemas reales de las ciencias e ingenierías tales como: Análisis estadístico, minería de datos, beneficio de una empresa, costo de producción de un determinado producto, entre otros. Así la importancia de la investigación radica en estudiar la generalización del método de máximo descenso para resolver problemas de optimización no necesariamente en el espacio euclidiano; en particular generalizaremos dicho método usando retracciones para variedades riemannianas, para tal fin debemos tener conocimientos básicos de geometría riemanniana, en específico de las variedades riemannianas.

CAPÍTULO II

MARCO TEÓRICO

1. Preliminares

En esta sección se presentará la simbología y notaciones a usar a lo largo de todo el trabajo, así como también, conceptos básicos de la teoría de variedades que serán utilizados en la siguiente sección como la base teórica para el desarrollo de nuestro método y lo complementaremos dando algunos ejemplos para su mejor entendimiento. Algunos de estos resultados solo serán referenciados ya que el material bibliográfico es extenso.

1.1. Simbología y notaciones

En este trabajo consideraremos las siguientes terminologías:

\mathcal{M}, \mathcal{N} : variedades.

x, y : puntos sobre una variedad.

ξ, η, ζ, χ : vectores tangentes.

$\xi_x, \eta_x, \zeta_x, \chi_x$: vectores tangentes en el punto x .

ψ, φ : cartas.

A, B : matrices cuadradas.

W, X, Y, Z : matrices.

$\mathcal{W}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$: subespacios lineales.

1.2. Nociones básicas de la teoría de variedades

Una variedad d -dimensional puede ser definido como un conjunto \mathcal{M} cubierto con una colección de pedazos coordinados, llamados cartas, que identifican ciertos subconjuntos de \mathcal{M} con subconjuntos abiertos de \mathbb{R}^d . La definición abstracta de una variedad esta basado en los conceptos de cartas y atlas; por tanto definiremos primero estos términos.

Definición 1.1 (Carta). Sea \mathcal{M} un conjunto, una carta d -dimensional del conjunto \mathcal{M} , es una biyección φ de un conjunto \mathcal{U} de \mathcal{M} sobre un subconjunto abierto de \mathbb{R}^d , es decir,

$$\begin{aligned} \varphi : \mathcal{U} \subset \mathcal{M} &\rightarrow \mathcal{S} \subset \mathbb{R}^d \\ x &\mapsto \varphi(x) \end{aligned}$$

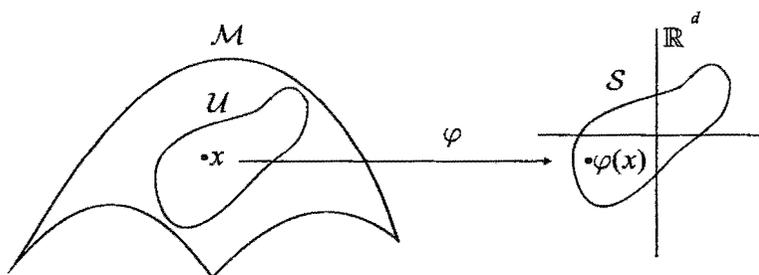


Figura 2.1: Carta

Notación: Una carta se denota por (\mathcal{U}, φ) y si no hay riesgo de confusión solo se escribe φ .

Observación 1.1. Dado una carta (\mathcal{U}, φ) y $x \in \mathcal{U}$, los elementos de $\varphi(x) \in \mathbb{R}^d$ son llamadas las coordenadas de x en la carta (\mathcal{U}, φ) .

Definición 1.2 (Atlas). Un atlas C^∞ de \mathcal{M} en \mathbb{R}^d es una colección de cartas $\{(\mathcal{U}_\alpha, \varphi_\alpha) : \alpha \in I\}$ del conjunto \mathcal{M} tal que:

1. $\bigcup_{\alpha \in I} \mathcal{U}_\alpha = \mathcal{M}$,
2. Para cualquier par $\alpha, \beta \in I$ con $\mathcal{U}_\alpha \cap \mathcal{U}_\beta \neq \emptyset$, los conjuntos $\varphi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$ y $\varphi_\beta(\mathcal{U}_\alpha \cap \mathcal{U}_\beta)$ son conjuntos abiertos en \mathbb{R}^d y el cambio de coordenadas

$$\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \subset \mathbb{R}^d \rightarrow \varphi_\beta(\mathcal{U}_\alpha \cap \mathcal{U}_\beta) \subset \mathbb{R}^d$$

es suave con una inversa suave (es decir, es un difeomorfismo).

Dos atlas \mathcal{A}_1 y \mathcal{A}_2 son equivalentes o compatibles si $\mathcal{A}_1 \cup \mathcal{A}_2$ es un atlas, en otras palabras para cada carta (\mathcal{U}, φ) en \mathcal{A}_2 , el conjunto de cartas $\mathcal{A}_1 \cup \{(\mathcal{U}, \varphi)\}$ es también un atlas.

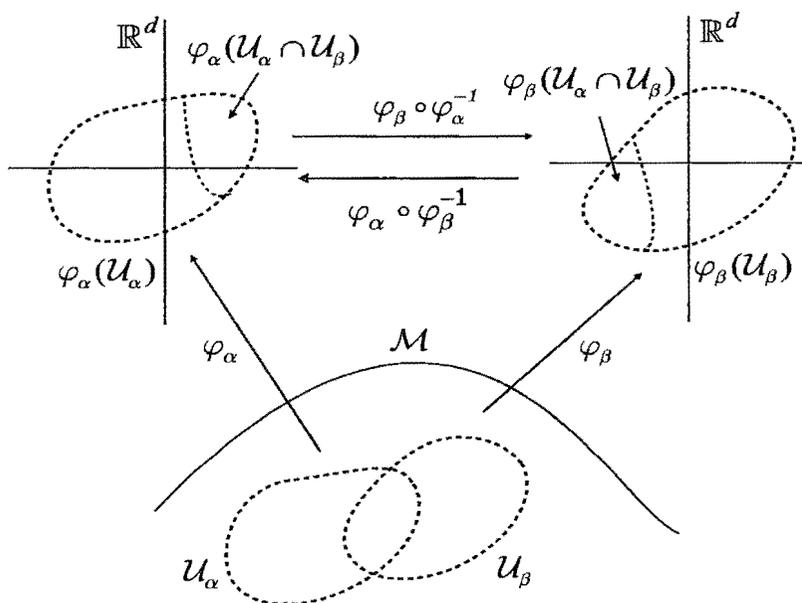


Figura 2.2: Cartas y cambio de coordenadas.

Dado un atlas \mathcal{A} , llamaremos \mathcal{A}^+ el conjunto de todas las cartas (U, φ) tal que $\mathcal{A} \cup \{(U, \varphi)\}$ es también un atlas; el atlas \mathcal{A}^+ es llamado atlas maximal o completo generado por el atlas \mathcal{A} .

A un atlas, esto es una colección de cartas $\{(U_\alpha, \varphi_\alpha)\}$ que cumple la condición 1 y 2 de la definición 1.2 y además es maximal, se le denomina estructura diferenciable.

Definición 1.3 (Variedad diferenciable). *Una variedad diferenciable (d -dimensional) es un par $(\mathcal{M}, \mathcal{A}^+)$, donde \mathcal{M} es un conjunto y \mathcal{A}^+ es un atlas maximal de \mathcal{M} en \mathbb{R}^d , tal que la topología inducida por \mathcal{A}^+ es de Hausdorff y segundo contable.*

Observaciones 1.2.

- *Un atlas maximal de un conjunto \mathcal{M} que induce una topología de Hausdorff y segundo contable es llamada una estructura de variedad sobre \mathcal{M} .*
- *Dado una variedad $(\mathcal{M}, \mathcal{A}^+)$, un atlas del conjunto \mathcal{M} cuyo atlas maximal es \mathcal{A}^+ es llamado un atlas de la variedad $(\mathcal{M}, \mathcal{A}^+)$; una carta del conjunto*

\mathcal{M} que pertenece a \mathcal{A}^+ es llamado una carta de la variedad $(\mathcal{M}, \mathcal{A}^+)$ y su dominio es un dominio coordinado de la variedad.

- Dado una carta φ sobre \mathcal{M} , la aplicación inversa φ^{-1} es llamada una parametrización local de \mathcal{M} . Una familia de parametrizaciones locales es equivalente a una familia de cartas.

1.3. Topología de una variedad

Antes de estudiar la topología de una variedad veamos algunas definiciones generales sobre topología.

Definición 1.4 (Topología). Una topología sobre un conjunto X es una colección τ de subconjuntos de X , llamados conjuntos abiertos tal que:

- (I) $X \wedge \phi \in \tau$.
- (II) La unión de elementos de cualquier subcolección de τ está en τ .
- (III) La intersección de elementos de cualquier subcolección finita de τ está en τ .

Un espacio topológico es un par (X, τ) donde X es un conjunto y τ es una topología en X .

Sea X un espacio topológico. Un subconjunto A de X se dice cerrado si el conjunto $X - A := \{x \in X : x \notin A\}$ es abierto.

Sea X un espacio topológico, X es T_1 , accesible o Frechet si para cualquier par de puntos distintos x e y de X , existe un conjunto abierto que contiene a x y no a y .

X es T_2 o Hausdorff si cualquier par de puntos distintos de X tienen vecindades disjuntas.

Una base para una topología en un conjunto X es una colección \mathcal{B} de subconjuntos de X tal que:

- i) Cada $x \in X$ pertenece a al menos un elemento de \mathcal{B} .
- ii) Si $x \in (B_1 \cap B_2)$ con $B_1, B_2 \in \mathcal{B}$, entonces existe $B_3 \in \mathcal{B}$ tal que $x \in B_3 \subseteq B_1 \cap B_2$.

Un espacio topológico X es llamado segundo contable si tiene una base contable para su topología.

Ya visto estas definiciones vamos a estudiar la topología de una variedad.

Como toda topología tiene una base, la base de la topología de un conjunto \mathcal{M} esta formado por la colección de dominios coordinados especificados por un atlas maximal \mathcal{A}^+ de un conjunto \mathcal{M} y llamaremos a esta topología, topología del atlas de \mathcal{M} inducida por \mathcal{A} .

En la topología del atlas un subconjunto \mathcal{V} de \mathcal{M} es abierto si y solo si para alguna carta (\mathcal{U}, φ) en \mathcal{A}^+ , $\varphi(\mathcal{V} \cap \mathcal{U})$ es un subconjunto abierto en \mathbb{R}^d . Equivalentemente un subconjunto \mathcal{V} de \mathcal{M} es abierto si y solo si para cada $x \in \mathcal{V}$, existe una carta (\mathcal{U}, φ) en \mathcal{A}^+ tal que $x \in \mathcal{U} \subset \mathcal{V}$.

Una topología de un atlas siempre satisface el axioma de separación T_1 , pero no todas las topologías de un atlas son de Hausdorff (T_2) pues dos puntos distintos no necesariamente tienen entornos distintos.

Para una variedad $(\mathcal{M}, \mathcal{A}^+)$ llamaremos topología de la variedad \mathcal{M} a la topología del atlas de \mathcal{M} inducida por \mathcal{A} .

A continuación a modo de ejemplo probaremos que los espacios vectoriales son variedades lineales.

Ejemplo 1.1 (Espacios vectoriales como variedades). *Sea \mathcal{E} un espacio vectorial d -dimensional. Entonces dado una base $(e_i)_{i=1, \dots, d}$ de \mathcal{E} , la función*

$$\begin{aligned} \psi : \mathcal{E} &\rightarrow \mathbb{R}^d \\ x &\mapsto \psi(x) = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} \end{aligned}$$

tal que $x = \sum_{i=1}^d x_i e_i$ es una carta del conjunto \mathcal{E} . Todas las cartas construidas de este modo forman una estructura diferenciable (atlas maximal) del conjunto \mathcal{E} , el cual dota a \mathcal{E} una estructura de variedad. Así cada espacio vectorial es una variedad lineal en una forma natural.

Prueba. Construyamos una estructura de variedad diferenciable en cualquier espacio vectorial real de dim d , para lo cual consideremos la base $e = (e_1, \dots, e_d)$ una base cualquiera de \mathcal{E} y consideremos la aplicación

$$\psi_1: \mathcal{E} \rightarrow \mathbb{R}^d$$

$$x \mapsto \psi_1(x) = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix},$$

donde $x = \sum_{i=1}^d x_i e_i$. Ahora veamos que esta aplicación es una carta del conjunto \mathcal{E} .

En efecto:

I) ψ_1 es inyectiva pues si,

$$\psi_1(x) = \psi_1(y) \Rightarrow \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_d \end{bmatrix} \Leftrightarrow x_i = y_i \forall i = 1, \dots, d \Leftrightarrow x = y$$

II) ψ_1 es sobreyectiva

Por demostrar que: $\forall z \in \mathbb{R}^d, \exists x \in \mathcal{E} / \psi_1(x) = z$

como $z \in \mathbb{R}^d \Rightarrow z = [z_1, \dots, z_d]$, luego sea $x = z_1 e_1 + \dots + z_d e_d \Rightarrow \psi_1(x) = z$

Como ψ_1 es inyectiva y sobreyectiva por tanto ψ_1 es biyectiva y por tanto es una carta que define una estructura diferenciable en el conjunto \mathcal{E} .

Ahora tenemos que probar que la estructura diferenciable es independiente de la base elegida, para esto tomemos otra base distinta $\tilde{e} = (\tilde{e}_1, \dots, \tilde{e}_d)$ y consideremos la aplicación

$$\psi_2: \mathcal{E} \rightarrow \mathbb{R}^d$$

$$x \mapsto \psi_2(x) = \begin{bmatrix} x_1 \\ \vdots \\ x_d \end{bmatrix},$$

donde $x = (x_1 \tilde{e}_1 + \dots + x_d \tilde{e}_d)$; de manera análoga a la anterior se prueba que esta aplicación también es una carta de \mathcal{E} que define una estructura diferenciable. Ahora como el cambio de base esta dado por $e_i = \sum a_{ij} \tilde{e}_j, i, j \in \{1, \dots, d\}$, entonces

el cambio de cartas estará determinado por, $\psi_2 \circ \psi_1^{-1} : \mathbb{R}^d \rightarrow \mathbb{R}^d$, definido como $\psi_2 \circ \psi_1^{-1}(r_1, \dots, r_d) = (\sum r_i a_{i1}, \dots, \sum r_i a_{id}), i \in \{1, \dots, d\}$. Puesto que cada componente de $\psi_2 \circ \psi_1^{-1}$ es de clase C^∞ , se tiene que los atlas $\{\psi_2\}$ y $\{\psi_1\}$ son compatibles, por tanto definen una estructura diferenciable llamada estructura diferenciable estándar de espacio vectorial real. \blacksquare

Definición 1.5 (Producto de variedades). *Sea \mathcal{M}_1 y \mathcal{M}_2 dos variedades de dimensión d_1 y d_2 respectivamente. El conjunto $\mathcal{M}_1 \times \mathcal{M}_2$ es definido como el conjunto de pares (x_1, x_2) , donde $x_1 \in \mathcal{M}_1$ y $x_2 \in \mathcal{M}_2$, es decir,*

$$\mathcal{M}_1 \times \mathcal{M}_2 = \{(x_1, x_2) / x_1 \in \mathcal{M}_1 \wedge x_2 \in \mathcal{M}_2\}$$

se llama producto de variedades.

Si $(\mathcal{U}_1, \varphi_1)$ y $(\mathcal{U}_2, \varphi_2)$ son cartas de la variedad \mathcal{M}_1 y \mathcal{M}_2 respectivamente, entonces la aplicación

$$\begin{aligned} \varphi_1 \times \varphi_2 : \mathcal{U}_1 \times \mathcal{U}_2 &\rightarrow \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \\ (x_1, x_2) &\mapsto (\varphi_1 \times \varphi_2)(x_1, x_2) = (\varphi_1(x_1), \varphi_2(x_2)) \end{aligned}$$

es una carta para el conjunto $\mathcal{M}_1 \times \mathcal{M}_2$. Todas las cartas así obtenidas forman un atlas para el conjunto $\mathcal{M}_1 \times \mathcal{M}_2$. Con la estructura diferenciable definida por este atlas, $\mathcal{M}_1 \times \mathcal{M}_2$ es llamada la variedad producto. Su topología es equivalente a la topología producto.

1.4. Funciones diferenciables

Sea F una función de una variedad \mathcal{M}_1 de dimensión d_1 en otra variedad \mathcal{M}_2 de dimensión d_2 . Sea x un punto de \mathcal{M}_1 . Eligiendo cartas φ_1 y φ_2 alrededor de x y $F(x)$ respectivamente, la función F alrededor de x puede ser estudiado por medio de las cartas, obteniendo la función

$$\hat{F} = \varphi_2 \circ F \circ \varphi_1^{-1} : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2} \tag{2.1}$$

llamada una representación coordenada de F .

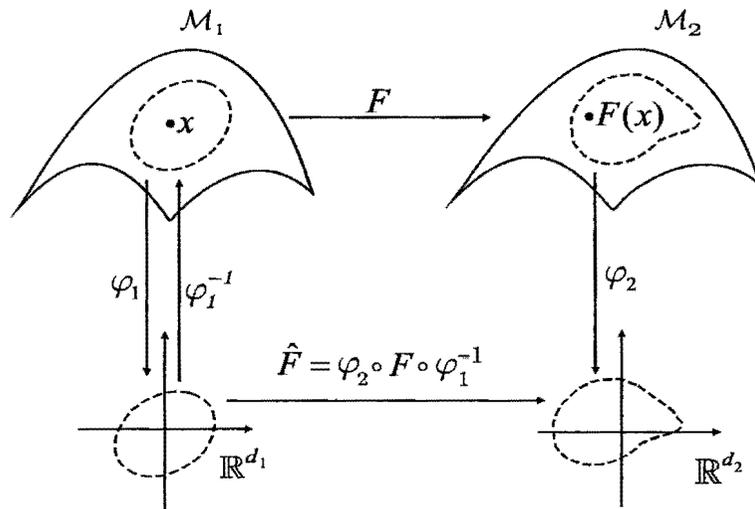


Figura 2.3: Funciones diferenciables

Decimos que F es diferenciable o suave en x si \hat{F} es de clase C^∞ en $\varphi_1(x)$. Una función $F : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ se dice suave, si es suave en cada punto de su dominio.

Definición 1.6 (Difeomorfismo (suave)). *Un difeomorfismo $F : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ es una biyección tal que F y su inversa F^{-1} son suaves.*

1.4.1. Inmersiones y submersiones

El concepto de inmersiones y submersiones harán posible para definir subvariedades y variedades cociente de forma concisa.

Sea $F : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ una función diferenciable de una variedad \mathcal{M}_1 de dimensión d_1 en otra variedad \mathcal{M}_2 de dimensión d_2 . Dado un punto x de \mathcal{M}_1 , el rango de F en x es la dimensión del rango de

$$D\hat{F}(\varphi_1(x))[\cdot] : \mathbb{R}^{d_1} \rightarrow \mathbb{R}^{d_2}$$

donde \hat{F} es una representación coordenada (2.1) de F alrededor de x , y $D\hat{F}(\varphi_1(x))$ denota el diferencial de \hat{F} en $\varphi_1(x)$.

La función F es llamada una **inmersión** si su rango es igual a d_1 en cada punto de su dominio ($d_1 \leq d_2$). Si su rango es igual a d_2 en cada punto de su dominio ($d_1 \geq d_2$), entonces esto es llamado una **submersión**.

Un punto $y \in \mathcal{M}_2$ se llama **valor regular** de F si el rango de F es d_2 en cada punto $x \in F^{-1}(y)$.

Ejemplo 1.2 (Ejemplo de inmersión). Para cada $n \geq m$ tenemos que $\mathbb{R}^n = \mathbb{R}^m \times \mathbb{R}^{n-m}$, entonces la aplicación inclusión inducida, definida por

$$\begin{aligned} \text{in} : \mathbb{R}^m &\rightarrow \mathbb{R}^m \times \mathbb{R}^{n-m} \\ r &\mapsto \text{in}(r) = (r, 0) \end{aligned}$$

es una inmersión. En efecto:

La matriz jacobiana de la inclusión es de la forma

$$J_{\text{in}}(r) = \begin{pmatrix} \text{id} \\ 0 \end{pmatrix}$$

y tiene rango m . Por lo tanto in es una inmersión.

Ejemplo 1.3 (Ejemplo de submersión). Para cada $m \geq n$ tenemos que $\mathbb{R}^m = \mathbb{R}^n \times \mathbb{R}^{m-n}$, entonces la aplicación proyección inducida, definida por

$$\begin{aligned} \text{pr}_1 : \mathbb{R}^n \times \mathbb{R}^{m-n} &\rightarrow \mathbb{R}^n \\ (r_1, r_2) &\mapsto \text{pr}_1(r_1, r_2) = r_1 \end{aligned}$$

es una submersión. En efecto:

La matriz jacobiana de pr_1 es:

$$J_{\text{pr}_1}(r_1, r_2) = (\text{id} \ 0)$$

que tiene rango n . Por lo tanto pr_1 es una submersión.

1.5. Subvariedades regulares

Un conjunto \mathcal{X} puede admitir muchas estructuras de variedad. Sin embargo si el conjunto \mathcal{X} es un subconjunto de una variedad $(\mathcal{M}, \mathcal{A}^+)$, entonces admite a lo más una estructura de variedad.

Sea $(\mathcal{M}, \mathcal{A}^+)$ y $(\mathcal{N}, \mathcal{B}^+)$ dos variedades tal que $\mathcal{N} \subset \mathcal{M}$. La variedad $(\mathcal{N}, \mathcal{B}^+)$ es llamada una **subvariedad inmersa** de $(\mathcal{M}, \mathcal{A}^+)$ si la aplicación inclusión $i : \mathcal{N} \rightarrow \mathcal{M}$ definida como $i(x) = x$ es una inmersión.

Sea $(\mathcal{N}, \mathcal{B}^+)$ una subvariedad de $(\mathcal{M}, \mathcal{A}^+)$. Ya que \mathcal{M} y \mathcal{N} son variedades son también espacios topológicos con la topología de la variedad. Si la topología de la variedad \mathcal{N} coincide con la topología del subespacio inducido por el espacio topológico \mathcal{M} , entonces \mathcal{N} es llamada una **subvariedad regular**, una subvariedad incrustada, encajada o simplemente una subvariedad de la variedad \mathcal{M} .

Proposición 1.1. *Sea \mathcal{N} un subconjunto de una variedad \mathcal{M} . Entonces \mathcal{N} admite a lo más una estructura diferenciable que lo hace una subvariedad regular. (Ver BOOTHBY W.M. (1986), [9], pag. 77.)*

Nota 1.1. *En la proposición anterior, la variedad \mathcal{M} es llamada el espacio encajante y si dicho espacio es el espacio $\mathbb{R}^{n \times p}$ o un subconjunto abierto de $\mathbb{R}^{n \times p}$, decimos que \mathcal{N} es una subvariedad matricial.*

Proposición 1.2 (Teorema de submersión). *Sea $F : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ una aplicación suave entre dos variedades de dimensión d_1 y d_2 , $d_1 > d_2$, y sea “ y ” un punto de \mathcal{M}_2 . Si y es un valor regular de F (es decir, el rango de F es igual a d_2 en cada punto de $F^{-1}(y)$) entonces $F^{-1}(y)$ es una subvariedad regular cerrada de \mathcal{M}_1 , y $\dim(F^{-1}(y)) = d_1 - d_2$. (Ver BOOTHBY W.M. (1986), [9], pag. 80.)*

Proposición 1.3 (Teorema de subinmersión). *Sea $F : \mathcal{M}_1 \rightarrow \mathcal{M}_2$ una aplicación suave entre dos variedades de dimensión d_1 y d_2 y sea y un punto de $F(\mathcal{M}_1)$. Si F tiene rango constante $k < d_1$ en una vecindad de $F^{-1}(y)$, entonces $F^{-1}(y)$ es una subvariedad regular cerrada de \mathcal{M}_1 de dimensión $d_1 - k$. (Ver BOOTHBY W.M. (1986), [9], pp. 79.)*

1.5.1. La variedad de Stiefel

Denotemos como $\text{St}(p, n)$, $p \leq n$ al conjunto de todas las matrices ortonormales es decir,

$$\text{St}(p, n) := \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\}, \quad (2.2)$$

donde I_p es la matriz identidad de orden $p \times p$. El conjunto $\text{St}(p, n)$ (con su estructura de subvariedad que se discutirá luego) es llamado la variedad de Stiefel (ortogonal o compacta).

A continuación veremos que $\text{St}(p, n)$ es una subvariedad regular de la variedad $\mathbb{R}^{n \times p}$. En efecto:

Claramente de la definición de $\text{St}(p, n)$ es un subconjunto del conjunto $\mathbb{R}^{n \times p}$; además el conjunto $\mathbb{R}^{n \times p}$ admite una estructura de variedad lineal. Ahora consideremos la función

$$\begin{aligned} F : \mathbb{R}^{n \times p} &\rightarrow S_{\text{sim}}(p) \\ X &\mapsto F(X) = X^T X - I_p, \end{aligned}$$

donde $S_{\text{sim}}(p)$ denota el conjunto de todas las matrices simétricas de orden $p \times p$. También se puede ver que $S_{\text{sim}}(p)$ es un espacio vectorial, pues como, $S_{\text{sim}}(p) \subseteq \mathbb{R}^{p \times p}$ probaremos que $aA + bB \in S_{\text{sim}}(p)$, para todo $A, B \in S_{\text{sim}}(p)$ y $a, b \in \mathbb{K}$, veamos:

$$A \in S_{\text{sim}}(p) \Rightarrow A = A^T \Rightarrow aA = aA^T$$

$$B \in S_{\text{sim}}(p) \Rightarrow B = B^T \Rightarrow bB = bB^T$$

entonces: $aA + bB = aA^T + bB^T = (aA)^T + (bB)^T = (aA + bB)^T$

por tanto, $aA + bB \in S_{\text{sim}}(p)$ y así $S_{\text{sim}}(p)$ es un espacio vectorial.

También: $\text{St}(p, n) = F^{-1}(0_p)$ pues:

$$\begin{aligned} F^{-1}(0_p) &= \{X \in \mathbb{R}^{n \times p} : f(X) = X^T X - I_p = 0_p\} \\ &= \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\} = \text{St}(p, n) \end{aligned}$$

Ahora para aplicar el teorema de submersión nos falta probar que 0_p es un valor regular, es decir, debemos probar que F es una submersión en cada punto X de $F^{-1}(0_p) = \text{St}(p, n)$; en efecto:

El hecho que el dominio de F es un espacio vectorial nos libra de tener que trabajar con cartas, solo necesitamos mostrar que:

Para todo $\hat{Z} \in S_{\text{sim}}(p)$, $\exists Z \in \mathbb{R}^{n \times p}$, tal que $D F(X)[Z] = \hat{Z}$; veamos:

$$D F(X)[Z] = \langle F'(X), Z \rangle = \langle X + X^T, Z \rangle = \langle X, Z \rangle + \langle X^T, Z \rangle = Z^T X + X^T Z,$$

ahora sea: $Z = \frac{1}{2} X \hat{Z}$, entonces;

$$D F(X)\left[\frac{1}{2} X \hat{Z}\right] = \frac{1}{2} \hat{Z}^T X^T X + \frac{1}{2} X^T X \hat{Z} = \hat{Z}, \text{ pues } X^T X = I_p \text{ y } \hat{Z}^T = \hat{Z},$$

por lo tanto, $\exists Z = \frac{1}{2}X\hat{Z}$ tal que $D F(X)[Z] = \hat{Z}$, esto quiere decir que el rango de la función $D F(X)[\cdot]$ es igual a $S_{\text{sim}}(p)$; por tanto el rango de F es igual a la dimensión de $S_{\text{sim}}(p)$ en cada punto de $\text{St}(p, n)$. Luego por el teorema de submersión se tiene que el conjunto $\text{St}(p, n)$ definido en (2.2) es una subvariedad regular de $\mathbb{R}^{n \times p}$.

Ahora para obtener la dimensión de $\text{St}(p, n)$, sabemos que el espacio vectorial $S_{\text{sim}}(p)$ tiene dimensión $\frac{1}{2}p(p+1)$, puesto que una matriz simétrica es completamente determinada por su parte triangular superior y también sabemos que la dimensión de $\mathbb{R}^{n \times p}$ es np ; por lo tanto por la proposición 1.2 tenemos que:

$$\dim \text{St}(p, n) = np - \frac{1}{2}p(p+1).$$

Puesto que $\text{St}(p, n)$ es una subvariedad regular de $\mathbb{R}^{n \times p}$, su topología es la topología del subconjunto inducido por $\mathbb{R}^{n \times p}$.

Observaciones 1.3.

- Para $p = 1$, la variedad de Stiefel $\text{St}(p, n)$ se reduce a la esfera unitaria S^{n-1} en \mathbb{R}^n .
- Para $p = n$, la variedad de Stiefel $\text{St}(p, n)$ se convierte en el grupo ortogonal O_n donde su dimensión es $\frac{1}{2}n(n-1)$.

1.6. Vectores tangentes y aplicaciones diferenciables

Existen muchos métodos para generalizar la noción de una derivada direccional

$$D f(x)[\eta] = \lim_{t \rightarrow 0} \frac{f(x + t\eta) - f(x)}{t} \quad (2.3)$$

de una función real f definida sobre una variedad. Uno de estos métodos para generalizar la derivada direccional es reemplazar $t \mapsto (x + t\eta)$ por una curva suave γ sobre \mathcal{M} por medio de x (es decir, $\gamma(0) = x$); por medio de esto se obtiene una derivada direccional bien definida $\left. \frac{d(f(\gamma(t)))}{dt} \right|_{t=0}$. Así tenemos una operación denotada por $\dot{\gamma}(0)$, que toma una función f , definida localmente en una vecindad de x , y devuelve el número real $\left. \frac{d(f(\gamma(t)))}{dt} \right|_{t=0}$.

1.6.1. Vectores tangentes

Sea \mathcal{M} una variedad. Una aplicación suave $\gamma : \mathbb{R} \rightarrow \mathcal{M}$ definida por $t \mapsto \gamma(t)$ se denomina una curva en \mathcal{M} . La idea de definir una derivada $\dot{\gamma}(t)$ como

$$\dot{\gamma}(t) := \lim_{\tau \rightarrow 0} \frac{\gamma(t + \tau) - \gamma(t)}{\tau} \quad (2.4)$$

requiere de una estructura de espacio vectorial para calcular la diferencia $\gamma(t + \tau) - \gamma(t)$, pues puede no estar definido en alguna variedad no lineal. Sin embargo dado una función real suave f sobre \mathcal{M} , la función $f \circ \gamma : t \mapsto f(\gamma(t))$ es una función de \mathbb{R} a \mathbb{R} con una derivada clásica bien definida. Esto será mejor explicado en la siguiente definición.

Definición 1.7. Sea x un punto en \mathcal{M} , sea γ una curva por medio de x en $t = 0$, y denotemos $\mathfrak{F}_x(\mathcal{M})$ al conjunto de funciones reales suaves definidas en una vecindad de x . La aplicación $\dot{\gamma}(0)$ definida por

$$\begin{aligned} \dot{\gamma}(0) : \mathfrak{F}_x(\mathcal{M}) &\rightarrow \mathbb{R} \\ f &\mapsto \dot{\gamma}(0)f := \left. \frac{d(f(\gamma(t)))}{dt} \right|_{t=0} \end{aligned} \quad (2.5)$$

es llamado el vector tangente a la curva γ en $t = 0$.

Nota 1.2. Cuando \mathcal{M} es un espacio vectorial \mathcal{E} , la aplicación $\dot{\gamma}(0)$ y la derivada $\dot{\gamma}'(0) := \lim_{t \rightarrow 0} \frac{1}{t}(\gamma(t) - \gamma(0))$ cumplen la siguiente relación

$$\dot{\gamma}f = D\bar{f}(\gamma(0))[\dot{\gamma}'(0)]$$

para todas las funciones \bar{f} definidas en una vecindad \mathcal{U} de $\gamma(0)$ en \mathcal{E} y donde f denota la restricción de \bar{f} a $\mathcal{U} \cap \mathcal{M}$.

A continuación definiremos formalmente la noción de vector tangente.

Definición 1.8 (Vector tangente). Un vector tangente ξ_x a una variedad en un punto x es una aplicación de $\mathfrak{F}_x(\mathcal{M})$ a \mathbb{R} tal que existe una curva γ sobre \mathcal{M} con $\gamma(0) = x$, satisfaciendo

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

para todo $f \in \mathfrak{F}_x(\mathcal{M})$. El punto x es llamado la base del vector tangente ξ_x .

Dado un vector tangente ξ a \mathcal{M} en x , existen infinitas curvas γ que realizan ξ (es decir, $\dot{\gamma}(0) = \xi$). Estas curvas pueden ser caracterizadas en coordenadas locales en la siguiente proposición.

Proposición 1.4. *Dos curvas γ_1 y γ_2 por medio de un punto x en $t = 0$ (es decir, $\gamma_1(0) = x = \gamma_2(0)$) satisfacen $\dot{\gamma}_1(0) = \dot{\gamma}_2(0)$ si y solo si, dado una carta (\mathcal{U}, φ) con $x \in \mathcal{U}$, se cumple que*

$$\left. \frac{d(\varphi(\gamma_1(t)))}{dt} \right|_{t=0} = \left. \frac{d(\varphi(\gamma_2(t)))}{dt} \right|_{t=0}$$

Prueba. Siendo (\mathcal{U}, φ) una carta, entonces,

$$\begin{aligned} \varphi : \mathcal{U} &\rightarrow \mathbb{R}^n \\ x &\mapsto \varphi(x) = (\varphi_1(x), \dots, \varphi_n(x)) \end{aligned}$$

Luego tenemos que cada $\varphi_i \in \mathfrak{F}_x(\mathcal{M})$, para todo $i = 1, \dots, n$ pues,

$$\begin{aligned} \varphi_i : \mathcal{U} &\rightarrow \mathbb{R} \\ x &\mapsto \varphi_i(x) \end{aligned}$$

Ahora si: $\dot{\gamma}_1(0) = \dot{\gamma}_2(0)$

$$\Rightarrow \dot{\gamma}_1(0)\varphi_i = \dot{\gamma}_2(0)\varphi_i, \forall i = 1, \dots, n$$

$$\Rightarrow \left. \frac{d(\varphi_i(\gamma_1(t)))}{dt} \right|_{t=0} = \left. \frac{d(\varphi_i(\gamma_2(t)))}{dt} \right|_{t=0}, \forall i = 1, \dots, n$$

$$\Rightarrow \left(\left. \frac{d(\varphi_1(\gamma_1(t)))}{dt} \right|_{t=0}, \dots, \left. \frac{d(\varphi_n(\gamma_1(t)))}{dt} \right|_{t=0} \right) = \left(\left. \frac{d(\varphi_1(\gamma_2(t)))}{dt} \right|_{t=0}, \dots, \left. \frac{d(\varphi_n(\gamma_2(t)))}{dt} \right|_{t=0} \right)$$

$$\Rightarrow \left. \frac{d(\varphi(\gamma_1(t)))}{dt} \right|_{t=0} = \left. \frac{d(\varphi(\gamma_2(t)))}{dt} \right|_{t=0}$$

Recíprocamente; siendo (\mathcal{U}, φ) una carta entonces $\varphi : \mathcal{U} \rightarrow \mathbb{R}^n$ es una biyección.

Ahora dado algún $f \in \mathfrak{F}_x(\mathcal{M})$ (arbitrario) tenemos:

$$\dot{\gamma}_1(0)f = \left. \frac{d(f(\gamma_1(t)))}{dt} \right|_{t=0} = \left. \frac{d((f \circ \varphi^{-1}) \circ (\varphi(\gamma_1(t))))}{dt} \right|_{t=0}$$

aplicando la regla de la cadena se tiene,

$$\begin{aligned}
 \dot{\gamma}_1(0)f &= [(f \circ \varphi^{-1})(\varphi(\gamma_1(0)))]' \frac{d(\varphi(\gamma_1(t)))}{dt} \Big|_{t=0} \\
 &= [(f \circ \varphi^{-1})(\varphi(\gamma_2(0)))]' \frac{d(\varphi(\gamma_2(t)))}{dt} \Big|_{t=0} \\
 &= \frac{d((f \circ \varphi^{-1})(\varphi(\gamma_2(t))))}{dt} \Big|_{t=0} \\
 &= \frac{d(f(\gamma_2(t)))}{dt} \Big|_{t=0} \\
 &= \dot{\gamma}_2(0)f
 \end{aligned}$$

Como f es arbitraria, entonces $\dot{\gamma}_1(0)f = \dot{\gamma}_2(0)f$, para todo $f \in \mathfrak{F}_x(\mathcal{M})$

Por lo tanto: $\dot{\gamma}_1(0) = \dot{\gamma}_2(0)$. ■

Definición 1.9. El espacio tangente a \mathcal{M} en x , denotado por $T_x\mathcal{M}$, es el conjunto de todos los vectores tangentes a \mathcal{M} en x , es decir,

$$T_x\mathcal{M} = \{\xi : \xi \text{ es un vector tangente a } \mathcal{M} \text{ en } x\}$$

Proposición 1.5. El espacio tangente $T_x\mathcal{M}$ admite una estructura de espacio vectorial.

Prueba. Dado $\dot{\gamma}_1(0), \dot{\gamma}_2(0) \in T_x(\mathcal{M})$ y $a, b \in \mathbb{R}$ definimos

$$(a\dot{\gamma}_1(0) + b\dot{\gamma}_2(0))f := a(\dot{\gamma}_1(0)f) + b(\dot{\gamma}_2(0)f).$$

Debemos probar que: $a\dot{\gamma}_1(0) + b\dot{\gamma}_2(0) \in T_x\mathcal{M}$, esto implica que debemos probar la existencia de una curva γ tal que $\dot{\gamma}(0) = a\dot{\gamma}_1(0) + b\dot{\gamma}_2(0)$.

En efecto:

Consideremos la carta (\mathcal{U}, φ) con $x \in \mathcal{U}$ y definimos la siguiente curva

$$\gamma(t) = \varphi^{-1}(a\varphi(\gamma_1(t)) + b\varphi(\gamma_2(t))),$$

entonces,

$$\varphi(\gamma(t)) = a\varphi(\gamma_1(t)) + b\varphi(\gamma_2(t)) \tag{2.6}$$

Luego:

$$\begin{aligned}
\dot{\gamma}(0)f &= \left. \frac{d(f(\gamma(t)))}{dt} \right|_{t=0}, \quad \forall f \in \mathfrak{F}_x(\mathcal{M}) \text{ (arbitraria)}. \\
&= \left. \frac{d((f \circ \varphi^{-1})(\varphi(\gamma(t))))}{dt} \right|_{t=0} \\
&= [(f \circ \varphi^{-1})(\varphi(\gamma(0)))]' \left. \frac{d\varphi(\gamma(t))}{dt} \right|_{t=0} \\
&= [(f \circ \varphi^{-1})(\varphi(\gamma(0)))]' \left. \frac{d(a\varphi(\gamma_1(t)) + b\varphi(\gamma_2(t)))}{dt} \right|_{t=0} \\
&= [(f \circ \varphi^{-1})(\varphi(\gamma(0)))]' \left\{ a \left. \frac{d(\varphi(\gamma_1(t)))}{dt} \right|_{t=0} + b \left. \frac{d(\varphi(\gamma_2(t)))}{dt} \right|_{t=0} \right\} \\
&= a[(f \circ \varphi^{-1})(\varphi(\gamma(0)))]' \left. \frac{d(\varphi(\gamma_1(t)))}{dt} \right|_{t=0} + b[(f \circ \varphi^{-1})(\varphi(\gamma(0)))]' \left. \frac{d(\varphi(\gamma_2(t)))}{dt} \right|_{t=0} \\
&= a \left. \frac{d((f \circ \varphi^{-1})(\varphi(\gamma_1(t))))}{dt} \right|_{t=0} + b \left. \frac{d((f \circ \varphi^{-1})(\varphi(\gamma_2(t))))}{dt} \right|_{t=0} \\
&= a \left. \frac{d(f(\gamma_1(t)))}{dt} \right|_{t=0} + b \left. \frac{d(f(\gamma_2(t)))}{dt} \right|_{t=0} \\
&= a\gamma_1(0)f + b\gamma_2(0)f \\
&= (a\dot{\gamma}_1(0) + b\dot{\gamma}_2(0))f,
\end{aligned}$$

donde la cuarta igualdad se obtiene por la ecuación (2.6) y la séptima igualdad se obtiene puesto que en la sexta igualdad se cumple que $\gamma(0) = \gamma_1(0) = \gamma_2(0)$.

Luego, como f es arbitraria, entonces se tiene que:

$$\dot{\gamma}(0) = (a\dot{\gamma}_1(0) + b\dot{\gamma}_2(0))f \quad \blacksquare$$

1.6.2. Vectores tangentes a un espacio vectorial

Sea \mathcal{E} un espacio vectorial y sea x un punto de \mathcal{E} . Un vector tangente ξ a \mathcal{E} en x es una aplicación

$$\begin{aligned}\xi : \mathfrak{F}_x(\mathcal{E}) &\rightarrow \mathbb{R} \\ f &\mapsto \xi f = \left. \frac{d(f(\gamma(t)))}{dt} \right|_{t=0},\end{aligned}$$

donde $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathcal{E}$ es una curva con $\gamma(0) = x$. Definiendo $\gamma'(0) \in \mathcal{E}$ como en (2.4) tenemos que

$$\xi f = Df(x)[\gamma'(0)]$$

Para mostrar esta igualdad usaremos la siguiente definición: Si f es una función real en un espacio euclidiano \mathcal{E} , entonces dado $x \in \mathcal{E}$, definimos $\text{grad } f(x)$ como el único elemento de \mathcal{E} que satisface

$$\langle \text{grad } f(x), h \rangle = Df(x)[h], \quad \forall h \in \mathcal{E}$$

En efecto:

$$\begin{aligned}\xi f &= \left. \frac{d(f(\gamma(t)))}{dt} \right|_{t=0} = [f(\gamma(0))]' \gamma'(0) \\ &= \langle \text{grad } f(\gamma(0)), \gamma'(0) \rangle; \quad \gamma(0) = x \in \mathcal{E} \\ &= Df(x)[\gamma'(0)]\end{aligned}$$

Además $\gamma'(0)$ no depende de la curva γ que realiza ξ . Esto define una correspondencia canónica lineal biyectiva

$$\begin{aligned}F : T_x \mathcal{E} &\rightarrow \mathcal{E} \\ \xi &\mapsto F(\xi) = \gamma'(0),\end{aligned}$$

la cual identifica a $T_x \mathcal{E}$ con \mathcal{E} , es decir:

$$T_x \mathcal{E} \simeq \mathcal{E}. \tag{2.7}$$

Para probar lo anterior usaremos el teorema de las transformaciones lineales que dice: Si A y B son espacios vectoriales y $T : A \rightarrow B$ una transformación lineal entonces $\frac{A}{\text{Nuc}(T)} \simeq \text{Im}(T)$

En efecto:

Como F es una correspondencia lineal biyectiva y $T_x\mathcal{E}$ y \mathcal{E} son espacios vectoriales, además F biyectiva, entonces, por teorema de álgebra lineal tenemos:

$$F \text{ inyectiva} \Rightarrow \text{Nuc}(F) = \{0\}$$

$$F \text{ sobreyectiva} \Rightarrow \text{Im}(F) = \mathcal{E}$$

Luego:

$$\begin{aligned} \frac{T_x\mathcal{E}}{\text{Nuc}(F)} &= \frac{T_x\mathcal{E}}{\{0\}} = \{\{0\} + \xi / \xi \in T_x\mathcal{E}\} \\ &= \{\xi / \xi \in T_x\mathcal{E}\} = T_x\mathcal{E} \end{aligned}$$

$$\text{Entonces, } T_x\mathcal{E} = \frac{T_x\mathcal{E}}{\text{Nuc}(F)} \simeq \text{Im}(F) = \mathcal{E}.$$

Por lo tanto, $T_x\mathcal{E} \simeq \mathcal{E}$.

1.6.3. Fibrado tangente

Dado una variedad \mathcal{M} , sea $T\mathcal{M}$ el conjunto de todos los vectores tangentes a \mathcal{M} :

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

Puesto que cada $\xi \in T\mathcal{M}$ esta en uno y solo un espacio tangente $T_x\mathcal{M}$, se deduce que \mathcal{M} es un cociente de $T\mathcal{M}$ con proyección natural,

$$\begin{aligned} \pi : T\mathcal{M} &\rightarrow \mathcal{M} \\ \xi \in T_x\mathcal{M} &\mapsto \pi(\xi) = x \end{aligned}$$

El conjunto $T\mathcal{M}$ admite una estructura de variedad llamada fibrado tangente de \mathcal{M} .

1.6.4. Campos vectoriales

Un campo vectorial ξ sobre una variedad \mathcal{M} es una función suave de \mathcal{M} al fibrado tangente $T\mathcal{M}$ que asigna a cada punto $x \in \mathcal{M}$ un vector tangente $\xi_x \in T_x\mathcal{M}$,

es decir,

$$\begin{aligned} f &: \mathcal{M} \rightarrow T\mathcal{M} \\ x &\mapsto f(x) = \xi_x \in T_x\mathcal{M} \end{aligned}$$

Dado un campo vectorial ξ sobre \mathcal{M} y una función real (suave) $f \in \mathfrak{F}(\mathcal{M})$, denotemos por ξf a la función real sobre \mathcal{M} definida por

$$(\xi f)(x) := \xi_x(f)$$

para todo $x \in \mathcal{M}$. La aplicación de dos campos vectoriales y la multiplicación de un campo vectorial por una función $f \in \mathfrak{F}(\mathcal{M})$ son definidas como:

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

Las operaciones anteriores preservan la suavidad. Denotamos por $\mathfrak{X}(\mathcal{M})$ al conjunto de campos vectoriales suaves dotados con las operaciones anteriores.

Sea (\mathcal{U}, φ) una carta de la variedad \mathcal{M} . El campo vectorial E_i sobre \mathcal{U} definida por

$$(E_i f)(x) := \partial_i(f \circ \varphi^{-1})(\varphi(x)) = D(f \circ \varphi^{-1})(\varphi(x))[e_i]$$

es llamado el i -ésimo campo vectorial coordenado de (\mathcal{U}, φ) . Estos campos vectoriales coordenados son suaves, y cada campo vectorial ξ admite la descomposición

$$\xi = \sum_i (\xi \varphi_i) E_i$$

sobre \mathcal{U} .

1.6.5. Diferencial de una aplicación

Sea $F : \mathcal{M} \rightarrow \mathcal{N}$ una aplicación suave entre dos variedades \mathcal{M} y \mathcal{N} . Sea ξ_x un vector tangente en un punto x de \mathcal{M} . La aplicación

$$\begin{aligned} D F(x) &: T_x\mathcal{M} \rightarrow T_{F(x)}\mathcal{N} \\ \xi &\mapsto D F(x)[\xi] \end{aligned}$$

es una aplicación lineal llamada la diferencial (función diferencial o derivada) de F en x .

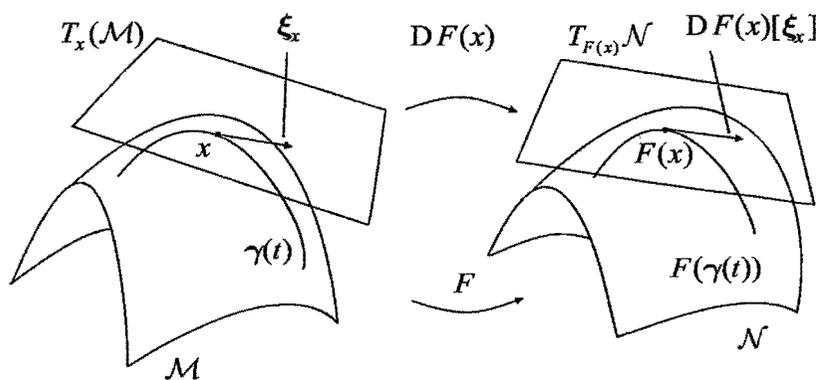


Figura 2.4: Aplicación diferencial de F en x

Notas 1.3.

1. La aplicación $DF(x)[\xi_x]$ de $\mathfrak{F}_{F(x)}(\mathcal{N})$ a \mathbb{R} definida por

$$(DF(x)[\xi])f := \xi(f \circ F) \quad (2.8)$$

es un vector tangente a \mathcal{N} en $F(x)$. El vector tangente $DF(x)[\xi_x]$ es realizado por $F \circ \gamma$, donde γ es alguna curva que realiza ξ_x .

2. Si \mathcal{M} y \mathcal{N} son variedades lineales, entonces con la identificación $T_x\mathcal{M} \simeq \mathcal{M}$ y $T_y\mathcal{N} \simeq \mathcal{N}$, $DF(x)$ se reduce a su definición clásica

$$DF(x)[\xi_x] = \lim_{t \rightarrow 0} \frac{F(x + t\xi_x) - F(x)}{t}. \quad (2.9)$$

1.6.6. Vectores tangentes a subvariedades regulares

Sea \mathcal{M} una subvariedad regular de un espacio vectorial \mathcal{E} y sea γ una curva en \mathcal{M} , con $\gamma(0) = x$. Definir

$$\dot{\gamma}(0) := \lim_{t \rightarrow 0} \frac{\gamma(t) - \gamma(0)}{t},$$

donde la expresión $\gamma(t) - \gamma(0)$ está bien definida puesto que $\gamma(t) \in \mathcal{E}$, para todo t . Así se sigue que $\dot{\gamma}(0)$ definida anteriormente es un elemento de $T_x\mathcal{E} \simeq \mathcal{E}$. Ya que γ es una curva en \mathcal{M} , también induce un vector tangente $\dot{\gamma}(0) \in T_x\mathcal{M}$.

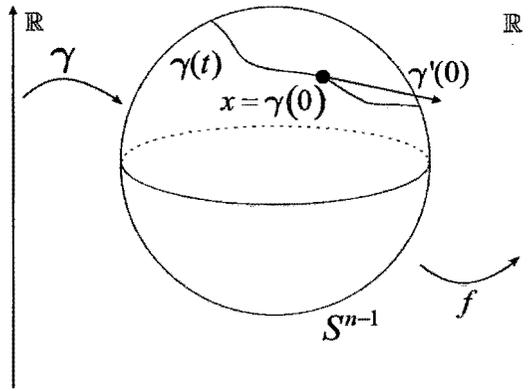


Figura 2.5: Curvas y vectores tangentes sobre la esfera S^{n-1}

También, $\dot{\gamma}(0)$ y $\dot{\gamma}(0)f$ están relacionados de la siguiente manera: Si \bar{f} es una función real en una vecindad \mathcal{U} de x en \mathcal{E} y f denota la restricción de \bar{f} a $\mathcal{U} \cap \mathcal{M}$, entonces tenemos

$$\dot{\gamma}(0)f = \frac{d}{dt} f(\gamma(t))|_{t=0} = \frac{d}{dt} \bar{f}(\gamma(t))|_{t=0} = D\bar{f}(x)[\dot{\gamma}(0)]. \quad (2.10)$$

Esto produce una identificación natural de $T_x\mathcal{M}$ con el conjunto

$$\{\dot{\gamma}(0) : \gamma \text{ es una curva en } \mathcal{M}, \gamma(0) = x\}, \quad (2.11)$$

el cual es un subespacio lineal del espacio vectorial $T_x\mathcal{E} \simeq \mathcal{E}$.

En vista de la identificación de $T_x\mathcal{M}$ con (2.11), se escribirá $\dot{\gamma}(t)$, $\dot{\gamma}'(t)$ y $\frac{d}{dt}\gamma(t)$ intercambiabilmente y también se usará la igualdad

$$T_x\mathcal{M} = \text{Nuc}(F(x)) \quad (2.12)$$

para denotar la identificación de $T_x\mathcal{M}$ con (2.11).

Cuando \mathcal{M} es definida como un conjunto nivel de una función de rango constante $F : \mathcal{E} \mapsto \mathbb{R}^n$, tenemos

$$T_x\mathcal{M} = \text{Nuc}(D F(x)), \quad (2.13)$$

esto quiere decir que el vector tangente a \mathcal{M} en x corresponde a los vectores ξ que satisfacen $D F(x)[\xi] = 0$

Ejemplo 1.4 (Espacio tangente a una esfera). Sea la curva $x : (-\varepsilon, \varepsilon) \rightarrow S^{n-1}$ definida como $t \mapsto x(t)$ en la esfera unitaria S^{n-1} que pasa por x_0 en $t = 0$ (es decir, $x(0) = x_0$). Ya que $x(t) \in S^{n-1}$ para todo t , tenemos

$$x^T(t) \cdot x(t) = 1, \forall t.$$

Diferenciando esta ecuación con respecto a t obtenemos:

$$\dot{x}^T(t)x(t) + x^T(t)\dot{x}(t) = 0,$$

$$\Rightarrow x^T(t)\dot{x}(t) = 0 \Rightarrow x^T(0)\dot{x}(0) = 0$$

$$\Rightarrow \dot{x}(0) \in V = \{z \in \mathbb{R}^n : x_0^T z = 0\}$$

entonces,

$$T_{x_0}S^{n-1} \subseteq V \tag{2.14}$$

Ahora sea $z \in V \Rightarrow x_0^T z = 0$. Luego consideremos la curva

$$\begin{aligned} x : (-\varepsilon, \varepsilon) &\rightarrow S^{n-1} \\ t &\mapsto x(t) = \frac{(x_0 + tz)}{\|x_0 + tz\|} \end{aligned}$$

Luego:

$$x'(t) = \frac{z}{\|z\|} = z; \text{ pues como } z \in S^{n-1} \Rightarrow \|z\| = 1$$

$$\Rightarrow \dot{x}(0) = z \Rightarrow z = \dot{x}(0) \in T_{x_0}S^{n-1}$$

entonces,

$$V \subseteq T_{x_0}S^{n-1} \tag{2.15}$$

Por tanto de (2.14) y (2.15) tenemos:

$$T_{x_0}S^{n-1} = \{z \in \mathbb{R}^n : x^T z = 0\},$$

el cual es el conjunto de todas los vectores ortogonales a x en \mathbb{R}^n .

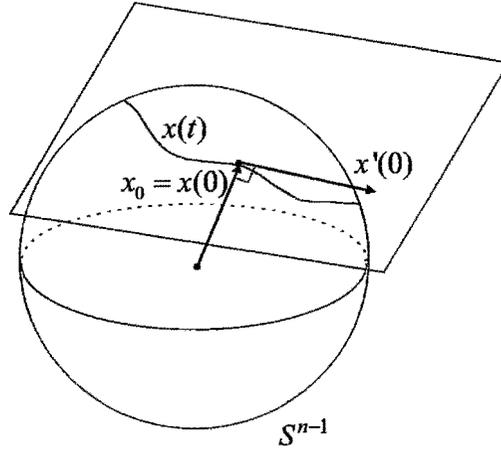


Figura 2.6: Espacio tangente sobre la esfera S^{n-1}

Ejemplo 1.5 (Variedad de Stiefel ortogonal).

Recordemos que $\text{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\}$. Usaremos la igualdad (2.12) para hallar $T_X \text{St}(p, n)$. Para esto consideremos la función

$$\begin{aligned} F : \mathbb{R}^{n \times p} &\rightarrow S_{\text{sim}}(p) \\ X &\mapsto F(x) = X^T X. \end{aligned}$$

Luego

$$F^{-1}(I_p) = \{X \in \mathbb{R}^{n \times p} : F(x) = I_p\} = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\} = \text{St}(p, n).$$

Probaremos que I_p es un valor regular de F , para esto probaremos que: Para todo $\hat{Z} \in S_{\text{sim}}(p)$, $\exists Z \in \mathbb{R}^{n \times p}$, tal que $DF(X)[Z] = \hat{Z}$. En efecto:

$$\begin{aligned} DF(x)[Z] &= \langle F'(x), Z \rangle = \langle X + X^T, Z \rangle \\ &= \langle X, Z \rangle + \langle X^T, Z \rangle = Z^T X + X^T Z. \end{aligned}$$

Ahora, sea $Z = \frac{1}{2} X \hat{Z}$, entonces,

$$DF(X)[(1/2)X\hat{Z}] = \frac{1}{2} \hat{Z}^T X^T X + \frac{1}{2} X^T X \hat{Z} = \hat{Z} \Rightarrow DF(X)[\mathbb{R}^{n \times p}] = S_{\text{sim}}(p).$$

Así el rango de F es igual a la dimensión de $S_{\text{sim}}(p)$, para todo $Z \in \text{dom } F$, en particular el rango de F es igual a la dimensión de $S_{\text{sim}}(p)$ para todo

$Z \in F^{-1}(I_p)$, por tanto I_p es un valor regular de F .

Por lo tanto:

$$T_x \text{St}(p, n) = \text{Nuc}(D F(x)) = \{Z \in \mathbb{R}^{n \times p} : Z^T X + X^T Z = 0\}$$

1.7. Métrica riemanniana, distancia y gradientes

Los vectores tangentes sobre variedades generaliza la noción de una derivada direccional. Para caracterizar que dirección de movimiento de x produce el mejor incremento en f , adicionalmente necesitamos una noción de longitud que aplica a vectores tangentes. Esto es hecho por dotar a cada espacio tangente $T_x \mathcal{M}$ con un producto interno $\langle \cdot, \cdot \rangle_x$, es decir, una forma bilineal, simétrica y definida positiva. El producto interno $\langle \cdot, \cdot \rangle_x$ induce una norma,

$$\|\xi_x\| := \sqrt{\langle \xi_x, \xi_x \rangle_x},$$

sobre $T_x \mathcal{M}$. La dirección de mayor ascenso es dado por

$$\arg \max_{\xi \in T_x \mathcal{M} : \|\xi_x\|=1} D f(x)[\xi_x]$$

Una variedad cuyos espacios tangentes con un producto interno variando suavemente es llamado una **variedad riemanniana**. El producto interno variando suavemente es llamado la **métrica riemanniana**. Se usarán las siguientes notaciones para denotar el producto interno de dos elementos ξ_x y ζ_x de $T_x \mathcal{M}$

$$g(\xi_x, \zeta_x) = g_x(\xi_x, \zeta_x) = \langle \xi_x, \zeta_x \rangle = \langle \xi_x, \zeta_x \rangle_x.$$

Estrictamente hablando, una variedad riemanniana es un par (\mathcal{M}, g) , donde \mathcal{M} es una variedad y g es una métrica riemanniana sobre \mathcal{M} .

Nota 1.4. *Un espacio vectorial dotado con un producto interno es una variedad riemanniana particular llamada espacio euclidiano.*

Sea (U, φ) una carta de una variedad riemanniana (\mathcal{M}, g) . Las componentes de g en la carta son dados por

$$g_{ij} := g(E_i, E_j),$$

donde, E_i denota el i -ésimo campo vectorial coordinado (ver sección 1.6.4). Así, para los campos vectoriales $\xi = \sum_i \xi^i E_i$ y $\zeta = \sum_i \zeta^i E_i$, tenemos

$$g(\xi, \zeta) = \langle \xi, \zeta \rangle = \sum_{i,j} g_{ij} \xi^i \zeta^j.$$

Note que las componentes g_{ij} son funciones reales sobre $\mathcal{U} \subseteq \mathcal{M}$. Podemos también definir las funciones reales $g_{ij} \circ \varphi^{-1}$ sobre $\varphi(\mathcal{U}) \subseteq \mathbb{R}^d$; usamos la misma notación para ambos. También usamos la notación $G : \hat{x} \mapsto G_{\hat{x}}$ para las funciones matriciales, tal que los elementos (i, j) de $G_{\hat{x}}$ son $g_{ij}|_{\hat{x}}$. Si denotamos por $\hat{\xi}_{\hat{x}} = D\varphi(\varphi^{-1}(\hat{x}))[\xi_x]$ y $\hat{\zeta}_{\hat{x}} = D\varphi(\varphi^{-1}(\hat{x}))[\zeta_x]$, con $\hat{x} = \varphi(x)$, la representación de ξ_x y ζ_x en la carta, entonces tenemos, en notación matricial,

$$g(\xi_x, \zeta_x) = \langle \xi_x, \zeta_x \rangle = \hat{\xi}_x^T G_{\hat{x}} \hat{\zeta}_x. \quad (2.16)$$

Se puede ver que G es una matriz simétrica y definida positiva en cada punto.

La **longitud de una curva** $\gamma : [a, b] \rightarrow \mathcal{M}$ sobre una variedad riemanniana (\mathcal{M}, g) es definida por

$$L(\gamma) = \int_a^b \sqrt{g(\dot{\gamma}(t), \dot{\gamma}(t))} dt.$$

La **distancia riemanniana** sobre una variedad riemanniana conectada (\mathcal{M}, g) es

$$\begin{aligned} \text{dist} : \mathcal{M} \times \mathcal{M} &\rightarrow \mathbb{R} \\ (x, y) &\mapsto \text{dist}(x, y) = \inf_{\Gamma} L(\gamma), \end{aligned}$$

donde, Γ es el conjunto de todas las curvas en \mathcal{M} uniendo puntos x e y .

Asumiendo que \mathcal{M} es de Hausdorff, se tiene que la distancia riemanniana define una métrica (ver BARRET O'NEILL (1983), [5]), es decir,

1. $\text{dist}(x, y) \geq 0$, con $\text{dist}(x, y) = 0 \Leftrightarrow x = y$ (definida positiva).
2. $\text{dist}(x, y) = \text{dist}(y, x)$ (simetría).
3. $\text{dist}(x, z) + \text{dist}(z, y) \geq \text{dist}(x, y)$ (desigualdad triangular).

Dado un campo escalar suave f sobre una variedad riemanniana \mathcal{M} , el gradiente de f en x denotado por $\text{grad } f(x)$, es definido como el único elemento de $T_x \mathcal{M}$ que satisface

$$\langle \text{grad } f(x), \xi \rangle_x = Df(x)[\xi], \quad \forall \xi \in T_x \mathcal{M}. \quad (2.17)$$

La expresión coordenada del grad f tiene la siguiente notación matricial,

$$\widehat{\text{grad}} f(\hat{x}) := G_{\hat{x}}^{-1} \text{Grad} \hat{f}(\hat{x}), \quad (2.18)$$

donde, G es la función matricial definida en (2.16) y Grad denota el gradiente euclidiano en \mathbb{R}^d ,

$$\text{Grad} \hat{f}(\hat{x}) := \begin{pmatrix} \partial_1 \hat{f}(\hat{x}) \\ \vdots \\ \partial_d \hat{f}(\hat{x}) \end{pmatrix}.$$

En realidad de (2.16) y (2.18), tenemos que

$$\langle \widehat{\text{grad}} f, \hat{\xi} \rangle = \hat{\xi}^T G(G^{-1} \text{Grad} \hat{f}) = \hat{\xi}^T \text{Grad} \hat{f} = D \hat{f}[\hat{\xi}] = D f[\xi].$$

El gradiente de una función tiene las siguientes propiedades notables de máximo ascenso:

- La dirección de $\text{grad} f(x)$ es la dirección de máximo ascenso de f en x :

$$\frac{\text{grad} f(x)}{\|\text{grad} f(x)\|} = \arg \max_{\xi \in T_x \mathcal{M}: \|\xi_x\|=1} D f(x)[\xi]$$

- La norma de $\text{grad} f(x)$ da la máxima pendiente de f en x :

$$\|\text{grad} f(x)\| = D f(x) \left[\frac{\text{grad} f(x)}{\|\text{grad} f(x)\|} \right]$$

Las dos propiedades son muy importantes en el ámbito de los métodos de optimización.

1.7.1. Subvariedades riemannianas

Sea \mathcal{M} una subvariedad regular de una variedad riemanniana $\overline{\mathcal{M}}$ desde que cada espacio tangente $T_x \mathcal{M}$ puede ser considerado como un subespacio de $T_x \overline{\mathcal{M}}$, la métrica riemanniana \overline{g} de $\overline{\mathcal{M}}$ induce una métrica riemanniana g sobre \mathcal{M} de acuerdo a

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

donde ξ y ζ del lado derecho de la igualdad son vistos como elementos de $T_x\overline{\mathcal{M}}$. Esto torna a \mathcal{M} en una variedad riemanniana de $\overline{\mathcal{M}}$.

El complemento ortogonal de $T_x\mathcal{M}$ en $T_x\overline{\mathcal{M}}$ es llamado el espacio normal a \mathcal{M} en x y es denotado por $(T_x\mathcal{M})^\perp$:

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

Algún elemento $\xi \in T_x\overline{\mathcal{M}}$ puede ser únicamente descompuesto en la suma de un elemento de $T_x\mathcal{M}$ y un elemento de $(T_x\mathcal{M})^\perp$:

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

donde, P_x denota la proyección ortogonal sobre $T_x\mathcal{M}$ y P_x^\perp denota la proyección ortogonal sobre $(T_x\mathcal{M})^\perp$.

Ejemplo 1.6. La esfera unitaria S^{n-1} es considerada una subvariedad riemanniana de \mathbb{R}^n , el producto interno es heredado del producto interno estándar en \mathbb{R}^n y es dado por

$$\langle \xi, \eta \rangle := \xi^T \eta \tag{2.19}$$

El espacio normal es

$$(T_x S^{n-1})^\perp = \{x\alpha : \alpha \in \mathbb{R}\},$$

y las proyecciones son dadas por

$$P_x\xi = (I - xx^T)\xi, \quad P_x^\perp\xi = xx^T\xi$$

para $x \in S^{n-1}$.

Observación 1.4. Sea \overline{f} una función objetivo definida sobre una variedad riemanniana $\overline{\mathcal{M}}$ y denotemos por f la restricción de \overline{f} a la subvariedad riemanniana \mathcal{M} . El gradiente de f es igual a la proyección del gradiente de \overline{f} sobre $T_x\mathcal{M}$.

$$\text{grad } f(x) = P_x \text{grad } \overline{f}(x). \tag{2.20}$$

En realidad, $P_x \text{grad } \bar{f}(x)$ pertenece a $T_x \mathcal{M}$ y la ecuación (2.17) es satisfecha, puesto que, $\forall \zeta \in T_x \mathcal{M}$ tenemos:

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

2. Método de máximo descenso usando retracciones en variedades riemannianas

En esta sección estudiaremos el método de máximo descenso en variedades riemannianas para resolver un problema de optimización de la forma:

$$\begin{cases} \text{mín } f(x) \\ \text{s.a.} \\ x \in \mathcal{M} \end{cases}$$

donde f es una función definida en una variedad riemanniana \mathcal{M} .

El método de máximo descenso genera una sucesión de puntos o fórmula de actualización

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

donde $\eta_k \in T_{x_k} \mathcal{M}$ ($T_{x_k} \mathcal{M}$ es el espacio tangente a la variedad \mathcal{M} en el punto x_k), t_k es un escalar para el cual se hará una búsqueda de Armijo para determinarlo y R es una aplicación llamada **retracción**, la cual estudiaremos a continuación.

2.1. Retracciones

Una retracción sobre una variedad \mathcal{M} es una aplicación suave R del fibrado tangente $T\mathcal{M}$ sobre \mathcal{M} , y si denotamos por R_x la restricción de R a $T_x \mathcal{M}$, tendremos:

$$\begin{aligned} R_x : T_x \mathcal{M} &\rightarrow \mathcal{M} \\ \xi &\mapsto R_x(\xi), \end{aligned}$$

Entonces, debe cumplir las siguientes propiedades:

(I) $R_x(0_x) = x$, donde 0_x denota el elemento cero de $T_x\mathcal{M}$.

(II) Con la identificación canónica $T_{0_x}T_x\mathcal{M} \simeq T_x\mathcal{M}$, R_x satisface

$$D R_x(0_x) = \text{id}_{T_x\mathcal{M}}, \quad (2.21)$$

donde $\text{id}_{T_x\mathcal{M}}$ denota la aplicación identidad sobre $T_x\mathcal{M}$.

Respecto a la condición (2.21) notemos que, puesto que R_x es una aplicación de $T_x\mathcal{M}$ a \mathcal{M} enviando 0_x a x , de esto se sigue que $D R_x(0_x)$ es una aplicación de $T_{0_x}(T_x\mathcal{M})$ a $T_x\mathcal{M}$ (ver sección 1.6.5). Ya que $T_x\mathcal{M}$ es un espacio vectorial, existe una identificación natural $T_{0_x}(T_x\mathcal{M}) \simeq T_x\mathcal{M}$ (ver sección 1.6.2). A la condición $D R_x(0_x) = \text{id}_{T_x\mathcal{M}}$ se llama la condición de rigidez local, que preserva gradientes en x .

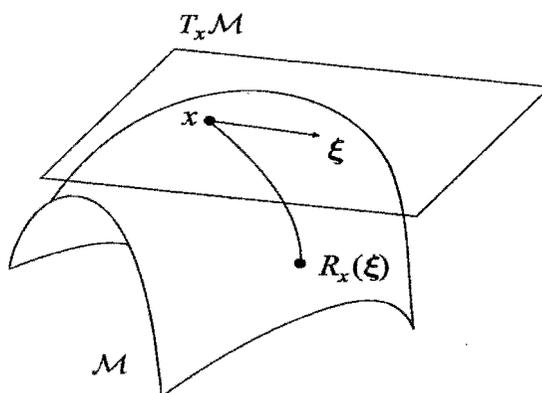


Figura 2.7: Retracción

Nota 2.1. Dado una función real f sobre una variedad \mathcal{M} dotada con una retracción R , denotemos la función \hat{f} como $\hat{f} = f \circ R$; para $x \in \mathcal{M}$ denotemos por

$$\hat{f}_x = f \circ R_x \quad (2.22)$$

a la restricción de \hat{f} a $T_x\mathcal{M}$. Note que \hat{f}_x es una función real sobre un espacio vectorial. Por la condición de rigidez local (2.21), tenemos (con la identificación canónica (2.7)) $D \hat{f}_x(0_x) = D f(x)$. Si \mathcal{M} es dotada con una métrica riemanniana

(y así $T_x\mathcal{M}$ con un producto interno), tenemos

$$\text{grad } \widehat{f}_x(0_x) = \text{grad } f(x) \quad (2.23)$$

A continuación haremos uso de muchas estructuras (subvariedades regulares) y objetos matemáticos (coordenadas locales, proyecciones, factorizaciones) que pueden ser aprovechadas para definir retracciones.

2.1.1. Retracciones sobre subvariedades regulares

Sea \mathcal{M} una subvariedad regular de un espacio vectorial \mathcal{E} . Recordemos que $T_x\mathcal{M}$ puede ser visto como un subespacio lineal de $T_x\mathcal{E}$ (ver sección 1.6.6) el cual puede ser identificado con \mathcal{E} (ver sección 1.6.2). Esto nos permite ligeramente abusar de notación, para considerar la suma $x + \xi$ de un punto x de \mathcal{M} , visto como un elemento de \mathcal{E} , y un vector tangente $\xi \in T_x\mathcal{M}$, visto como un elemento de $T_x\mathcal{E} \simeq \mathcal{E}$. En este sentido es conveniente definir una retracción de la siguiente manera. Dado $x \in \mathcal{M}$ y $\xi \in T_x\mathcal{M}$, calcular $R_x(\xi)$ como:

1. Movimiento a lo largo de ξ para obtener el punto $x + \xi$ en el espacio encajante lineal.
2. Proyectando el punto $x + \xi$ sobre la variedad \mathcal{M} .

El asunto ahora es definir una proyección que torne el procedimiento en una retracción bien definida y que sea computacionalmente eficiente, para esto nos basaremos en la descomposición de matrices tales como la descomposición que incluye la factorización QR y la descomposición polar. El propósito de esta sección es desarrollar la teoría de retracciones basadas en descomposición.

Sea \mathcal{M} una variedad regular de un espacio vectorial \mathcal{E} y sea \mathcal{N} una variedad abstracta tal que $\dim(\mathcal{M}) + \dim(\mathcal{N}) = \dim(\mathcal{E})$. Suponer que existe un difeomorfismo

$$\begin{aligned} \phi : \mathcal{M} \times \mathcal{N} &\rightarrow \mathcal{E}_* \\ (F, G) &\mapsto \phi(F, G) \end{aligned}$$

donde \mathcal{E}_* es un subconjunto abierto de \mathcal{E} (así \mathcal{E}_* es una subvariedad abierta de \mathcal{E}), con un elemento neutral $I \in \mathcal{N}$ satisfaciendo

$$\phi(F, I) = F, F \in \mathcal{M}.$$

Proposición 2.1. *Bajo las suposiciones dadas anteriormente sobre ϕ , la aplicación*

$$R_X(\mathcal{E}) := \pi_1(\phi^{-1}(X + \xi));$$

donde, $\pi_1 : \mathcal{M} \times \mathcal{N} \rightarrow \mathcal{M}$ definida como $\pi_1(F, G) = F$ es la proyección sobre la primera componente; define una retracción sobre \mathcal{M} .

Prueba. Puesto que \mathcal{E}_* es abierto, se sigue que $X + \xi \in \mathcal{E}_*$ para todo ξ en algún entorno de 0_X , ahora como ϕ es difeomorfismo, entonces ϕ^{-1} está definido sobre todo el espacio \mathcal{E}_* (es decir, $\text{dom}(\phi^{-1}) = \mathcal{E}_*$), ahora como $\pi_1 \circ \phi^{-1} : \mathcal{E}_* \rightarrow \mathcal{M}$ entonces se tiene que $R_X(\xi)$ está definido para todo ξ en un entorno del origen de $T_X\mathcal{M}$.

Veamos que la aplicación R es suave; como ϕ es difeomorfismo entonces ϕ y ϕ^{-1} son diferenciables, y se sabe que la proyección es diferenciable, por tanto $R_X(\xi) := \pi_1(\phi^{-1}(X + \xi))$ es diferenciable y por tanto también es continua.

Ahora probemos las condiciones (I) y (II) de la definición de retracción.

(I) Por demostrar que: $R_X(0_X) = X$, en efecto:

$$R_X(0_X) = \pi_1(\phi^{-1}(X + 0_X)) = \pi_1(\phi^{-1}(X)) = \pi_1(X, I) = X$$

(II) Propiedad de rigidez local, para esto notemos que $\forall \xi \in T_X\mathcal{M}$, tenemos,

$$D_1 \phi(X, I)[\xi] = \left\langle \frac{\partial \phi(X, I)}{\partial X}, \xi \right\rangle = \langle 1, \xi \rangle = \xi$$

También:

$$\begin{aligned} D \phi(X, I)[(\xi, 0)] &= \langle \text{grad } \phi(X, I), (\xi, 0) \rangle \\ &= \langle \text{grad } X, (\xi, 0) \rangle \\ &= \langle (1, 0), (\xi, 0) \rangle \\ &= \xi \end{aligned}$$

Por tanto:

$$D_1 \phi(X, I)[\xi] = D \phi(X, I)[(\xi, 0)] = \xi$$

Ahora ya que: $\pi_1 \circ \phi^{-1}(\phi(F, I)) = F$, se tiene que, para todo $\xi \in T_X \mathcal{M}$,

$$D(\pi_1 \circ \phi^{-1})(\phi(X, I))[D_1 \phi(X, I)[\xi]] = \xi,$$

pues $D X[\xi] = \langle \text{grad } X, \xi \rangle = \langle 1, \xi \rangle = \xi$. Entonces:

$$\begin{aligned} \xi &= D(\pi_1 \circ \phi^{-1})(\phi(X, I))[D_1 \phi(X, I)[\xi]] \\ &= D(\pi_1 \circ \phi^{-1})(X)[\xi] = D R_X(0_X)[\xi] \end{aligned}$$

$$\Rightarrow D R_X(0_X) = \text{id}_{T_X \mathcal{M}}.$$

Por lo tanto, R_X es una retracción. ■

Ejemplo 2.1 (Retracción sobre el grupo ortogonal). Sea $\mathcal{M} = O_n = \text{St}(n, n)$ el grupo ortogonal. La descomposición QR de una matriz $A \in \mathbb{R}_*^{n \times n}$ es la descomposición de A como $A=QR$, donde Q pertenece a O_n y $R \in S_{\text{sup}+}(n)$ (el conjunto de todas las matrices triangulares superiores con elementos de la diagonal estrictamente positivos). Luego la inversa de la descomposición QR es la aplicación

$$\begin{aligned} \phi : O_n \times S_{\text{sup}+}(n) &\rightarrow \mathbb{R}_*^{n \times n} \\ (Q, R) &\mapsto QR. \end{aligned}$$

Ahora denotemos por $\text{qf} := \pi_1 \circ \phi^{-1}$ a la aplicación que envía una matriz al factor Q de su descomposición QR . La aplicación qf puede ser calculada usando la ortonormalización de Gram-Schmidt.

Verifiquemos que ϕ satisface todas las suposiciones de la Proposición 2.1. En efecto:

La matriz identidad será el elemento neutral pues $\phi(Q, I) = QI = Q$, para todo $Q \in O_n$, además por la existencia y unicidad de la descomposición se tiene que ϕ es biyectiva, también como ϕ es la restricción de una función suave (producto de matrices) entonces ϕ es suave. Veamos que ϕ^{-1} es también suave, para esto notemos que la matriz Q es obtenida por el proceso de Gram-Schmidt, el cual es C^∞ sobre

el conjunto de matrices de rango completo. Ahora como la segunda componente R es obtenida como $Q^{-1}M$, se deduce ϕ^{-1} es C^∞ . En conclusión las suposiciones de la Proposición 2.1 se cumplen, por lo tanto una retracción para el grupo ortogonal será,

$$R_X(X\Omega) := \text{qf}(X + X\Omega) = \text{qf}(X(I + \Omega)) = X \text{qf}(I + \Omega).$$

Ejemplo 2.2 (Retracción sobre la variedad de Stiefel). Considerar la variedad de Stiefel $\text{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\}$, tal como en el caso del grupo ortogonal, la retracción será,

$$R_X(\xi) := \text{qf}(X + \xi),$$

donde $\text{qf}(A)$ denota el factor Q de la descomposición de $A \in \mathbb{R}_*^{n \times p}$; como $A = QR$, donde $Q \in \text{St}(p, n)$ y R es una matriz triangular superior de orden $p \times p$ con elementos de su diagonal estrictamente positivos.

Ejemplo 2.3 (Retracción sobre la esfera S^{n-1}).

Sean $\mathcal{M} = S^{n-1}$ y $\mathcal{N} = \{x \in \mathbb{R} : x > 0\}$, y considerar la aplicación

$$\begin{aligned} \phi : \mathcal{M} \times \mathcal{N} &\rightarrow \mathbb{R}_*^n \\ (x, r) &\mapsto xr. \end{aligned}$$

Es sencillo de verificar que ϕ es un difeomorfismo; luego por la Proposición 2.1 se obtiene la retracción

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

definida para todo $\xi \in T_x S^{n-1}$. El punto $R_x(\xi)$ es el punto de S^{n-1} que minimiza la distancia a $x + \xi$.

2.2. Método del máximo descenso

Este método de búsqueda lineal sobre variedades riemannianas está basado en la fórmula de actualización

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

donde η_k está en $T_{x_k} \mathcal{M}$, t_k es un escalar y R_{x_k} es una aplicación retracción. Para obtener los resultados de convergencia estudiaremos algunas restricciones que deben ser impuestas a η_k y t_k .

Definición 2.1 (Sucesión relacionada con el gradiente). *Dado una función objetivo f sobre una variedad riemanniana \mathcal{M} , una sucesión $\{\eta_k\}$, $\eta_k \in T_{x_k}\mathcal{M}$, es relacionada al gradiente si, para alguna subsucesión $\{x_k\}_{k \in \mathcal{K}}$ de $\{x_k\}$ que converge a un punto no crítico de f , la subsucesión correspondiente $\{\eta_k\}_{k \in \mathcal{K}}$ es acotada y satisface*

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

Definición 2.2 (Punto de Armijo). *Dado una función objetivo f sobre una variedad riemanniana \mathcal{M} con retracción R , un punto $x \in \mathcal{M}$, un vector tangente $\eta \in T_x\mathcal{M}$, y escalares $\bar{\alpha} > 0$, $\beta, \sigma \in (0, 1)$, el vector de Armijo es $\eta^A = t^A\eta = \beta^m\bar{\alpha}\eta$, donde m es el menor entero no negativo tal que*

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

El real t^A es el tamaño de paso de Armijo.

A continuación proponemos el algoritmo 1 que describe una búsqueda lineal acelerada.

Algoritmo 1 Búsqueda lineal acelerada

Entrada: Una variedad riemanniana \mathcal{M} , una función objetivo f continuamente diferenciable sobre \mathcal{M} , una retracción R de $T\mathcal{M}$ a \mathcal{M} , escalares $\bar{\alpha} > 0; c, \beta, \sigma \in (0, 1)$ y una iteración inicial $x_0 \in \mathcal{M}$.

Salida: Sucesión de iteraciones $\{x_k\}$.

- 1: **para** $k = 0, 1, 2, \dots$ **hacer**
- 2: Escoger $\eta_k \in T_{x_k}\mathcal{M}$ tal que la sucesión $\{\eta_k\}_{k=0,1,\dots}$ es relacionada al gradiente.
- 3: Seleccionar x_{k+1} tal que

$$f(x_k) - f(x_{k+1}) \geq c(f(x_k) - f(R_{x_k}(t_k^A\eta_k))), \quad (2.24)$$

donde, t_k^A es el tamaño de paso de Armijo para $\bar{\alpha}, \beta, \sigma$ y η_k dados.

- 4: **fin de para**
-

El algoritmo 1 esta dado de manera general pues se pueden acoplar muchos métodos de interés mientras satisfagan ciertas propiedades fundamentales de convergencia. En particular la elección de $x_{k+1} = R_{x_k}(t_k^A \eta_k)$ en el paso 3 del algoritmo 1 satisface (2.24), pero esta elección no es única, por ejemplo la elección de $x_{k+1} = R_{x_k}(t_k^* \eta_k)$, donde $t_k^* = \arg \min_t f(R_{x_k}(t \eta_k))$, también satisface (2.24), esta elección es conocida como búsqueda lineal exacta.

2.3. Análisis de convergencia

Primero definiremos las nociones de convergencia y puntos límite sobre variedades, luego daremos un resultado de convergencia global para el algoritmo 1.

2.3.1. Convergencia sobre variedades

La noción de convergencia sobre variedades es la generalización del caso en \mathbb{R}^n , veamos.

Definición 2.3. *Una sucesión infinita $\{x_k\}_{k=0,1,\dots}$ de puntos de una variedad \mathcal{M} se dice que es convergente si existe una carta (\mathcal{U}, ψ) de \mathcal{M} , un punto $x_* \in \mathcal{U}$, y $k > 0$ tal que x_k esta en \mathcal{U} para todo $k \geq K$ y tal que la sucesión $\{\psi(x_k)\}_{k=K, K+1, \dots}$ converge a $\psi(x_*)$.*

El punto $\psi^{-1}(\lim_{k \rightarrow \infty} \psi(x_k))$ es llamado el límite de la sucesión convergente $\{x_k\}_{k=0,1,\dots}$. Cada sucesión convergente de una variedad (Hausdorff) tiene uno y solo un punto límite. Esto se justifica dado que la variedad es de Hausdorff y se sabe que en un espacio de Hausdorff toda sucesión convergente tiene un límite único. (Para una demostración de lo mencionado, ver SEYMOUR LIPSCHUTZ (1965), [21], pag. 145.)

Nota 2.2. *Una definición equivalente y más concisa de convergencia es la siguiente: Una sucesión sobre una variedad es convergente si esta es convergente en la topología de la variedad, es decir, existe un punto x_* tal que cada entorno de x_* contiene todos los términos de la sucesión, excepto un número finito.*

Definición 2.4. Dado una sucesión $\{x_k\}_{k=0,1,\dots}$ decimos que x es un punto de acumulación o punto límite de la sucesión, si existe una subsucesión $\{x_{j_k}\}_{k=0,1,\dots}$ que converge a x . El conjunto de puntos de acumulación de una sucesión es llamada el conjunto límite de la sucesión.

2.3.2. Convergencia de los métodos de búsqueda lineal

A continuación daremos un resultado de convergencia para el método de búsqueda lineal definido en el algoritmo 1, el cual es una generalización de la teoría clásica en \mathbb{R}^n .

Teorema 2.1. Sea $\{x_k\}$ una sucesión infinita de iteraciones generada por el algoritmo 1. Entonces cada punto de acumulación de $\{x_k\}$ es un punto crítico de la función objetivo f .

Demostración.

Procederemos la prueba por contradicción, supongamos que existe una subsucesión $\{x_k\}_{k \in \mathcal{K}}$ convergiendo a x_* con $\text{grad } f(x_*) \neq 0$. Puesto que $\{f(x_k)\}$ es no creciente, se deduce que la sucesión $\{f(x_k)\}$ converge a $f(x_*)$. Así $f(x_k) - f(x_{k+1})$ tiende a cero. Ahora por construcción del algoritmo tenemos,

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

donde $\alpha_k \eta_k$ es el vector de Armijo. Puesto que $\{\eta_k\}$ es gradiente relacionado, se tiene que $\{\alpha_k\}_{k \in \mathcal{K}} \rightarrow 0$. De esto se deduce que para todo k más grande que algún \bar{k} , $\alpha_k < \bar{\alpha}$, lo cual significa que $\alpha_k = \beta^m \bar{\alpha}$ para algún $m \geq 1$ y esto implica que el tamaño de paso previamente intentado $\beta^{m-1} \bar{\alpha} = \alpha_k / \beta$ no satisface la condición de Armijo. En otras palabras,

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

Denotando por,

$$\tilde{\eta}_k = \frac{\eta_k}{\|\eta_k\|} \quad y \quad \tilde{\alpha}_k = \frac{\alpha_k \|\eta_k\|}{\beta}, \quad (2.25)$$

La desigualdad anterior es igual a,

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

donde \hat{f} es definida como en (2.3). Luego por el teorema del valor medio se tiene

$$-\left\langle \text{grad } \hat{f}_{x_k}(t\tilde{\eta}_k) \right\rangle_{x_k} < -\sigma \langle \text{grad } f(x_k), \eta_k \rangle_{x_k}, \quad \forall k \in \mathcal{K}, k \geq \bar{k}, \quad (2.26)$$

donde t pertenece al intervalo $[0, \tilde{\alpha}_k]$. Puesto que $\{\alpha_k\}_{k \in \mathcal{K}} \rightarrow 0$ y como η_k es gradiente relacionado, por tanto acotado, se tiene que $\{\tilde{\alpha}_k\}_{k \in \mathcal{K}} \rightarrow 0$. Además, como $\tilde{\eta}_k$ tiene norma unitaria y su base x_k converge sobre el conjunto de índices \mathcal{K} , se tiene que $\{\eta_k\}_{k \in \mathcal{K}}$ está incluido en algún subconjunto compacto del fibrado tangente $T\mathcal{M}$, y por lo tanto existe un conjunto de índices $\tilde{\mathcal{K}} \subseteq \mathcal{K}$ tal que $\{\tilde{\eta}_k\}_{k \in \tilde{\mathcal{K}}} \rightarrow \tilde{\eta}_*$ para algún $\tilde{\eta}_* \in T_{x_*}\mathcal{M}$ con $\|\tilde{\eta}_*\| = 1$. Ahora tomamos límite en (2.26) sobre $\tilde{\mathcal{K}}$. Como la métrica riemanniana es continua (por definición), y $f \in C^1$, y $\text{grad } f_{x_k}(0) = \text{grad } f(x_k)$, se obtiene,

$$-\langle \text{grad } f(x_*), \tilde{\eta}_* \rangle_{x_*} \leq -\sigma \langle \text{grad } f(x_*), \tilde{\eta}_* \rangle_{x_*},$$

y puesto que $0 < \sigma < 1$, se tiene que $\langle \text{grad } f(x_*), \tilde{\eta}_* \rangle_{x_*} \geq 0$. De otro lado, de el hecho que $\{\eta_k\}$ es gradiente relacionado, obtenemos que $\langle \text{grad } f(x_*), \tilde{\eta}_* \rangle < 0$, lo cual es una contradicción. \blacksquare

Corolario 2.1. *Sea $\{x_k\}$ una sucesión infinita de iteraciones generada por el algoritmo 1. Suponer que el conjunto de nivel $\mathcal{L} = \{x \in \mathcal{M} : f(x) \leq f(x_0)\}$ es compacto (el cual se cumple en particular cuando el mismo \mathcal{M} es compacto). Entonces $\lim_{k \rightarrow \infty} \|\text{grad } f(x_k)\| = 0$.*

Demostración. Por contradicción, supongamos lo contrario. Entonces existe una subsucesión $\{x_k\}_{k \in \mathcal{K}}$ y $\varepsilon > 0$ tal que $\|\text{grad } f(x_k)\| > \varepsilon$ para todo $k \in \mathcal{K}$. Como f es no creciente sobre $\{x_k\}$, se tiene que $x_k \in \mathcal{L}$ para todo k . También como \mathcal{L} es compacto, $\{x_k\}_{k \in \mathcal{K}}$ tiene un punto de acumulación x_* en \mathcal{L} . Por la continuidad de $\text{grad } f$, tenemos que $\|\text{grad } f(x_*)\| \geq \varepsilon$; es decir, x_* no es un punto crítico, lo cual contradice al teorema 2.1. \blacksquare

2.4. Estabilidad de puntos fijos

El Teorema 2.1 afirma que solo los puntos críticos de una función objetivo f pueden ser puntos de acumulación de las sucesiones $\{x_k\}$ generadas por el algoritmo 1. Este resultado nos da información útil sobre el comportamiento del algoritmo 1, pero no nos especifica si dichos puntos de acumulación son minimizadores locales, maximizadores locales o puntos silla.

Nota 2.3. *En la práctica, si un punto inicial x_0 es cuidadosamente escogido, el algoritmo 1 genera sucesiones cuyos puntos de acumulación son mínimos locales de la función objetivo. Para sostener dicha observación haremos el siguiente análisis de estabilidad de puntos críticos.*

Definiciones 2.5.

1. Sea F una aplicación de \mathcal{M} a \mathcal{M} . Un punto $x_* \in \mathcal{M}$ es un punto fijo de F si $F(x_*) = x_*$. Denotemos como $F^{(n)}$ al resultado de n aplicaciones de F en x , es decir,

$$F^{(1)}(x) = F(x), \quad F^{(i+1)}(x) = F(F^{(i)}(x)), \quad i = 1, 2, \dots$$

2. Un punto fijo x_* de F es un punto estable de F si, para cada vecindad \mathcal{U} de x_* , existe una vecindad \mathcal{V} de x_* tal que, para todo $x \in \mathcal{V}$ y para todo entero positivo n , se cumple que $F^{(n)}(x) \in \mathcal{U}$.
3. El punto fijo x_* es asintóticamente estable si este es estable y además $\lim_{n \rightarrow \infty} F^{(n)}(x) = x_*$ para todo x suficientemente cerca a x_* .
4. El punto fijo x_* es inestable si este no es estable; en otras palabras, existe un entorno \mathcal{U} de x_* tal que, para todo entorno \mathcal{V} de x_* , existe un punto $x \in \mathcal{V}$ tal que $F^{(n)}(x) \notin \mathcal{U}$ para algún n .
5. Decimos que F es una aplicación descenso para una función objetivo f si,

$$f(F(x)) \leq f(x), \quad \forall x \in \mathcal{M}.$$

Teorema 2.2 (Puntos fijos inestables). *Sea $F : \mathcal{M} \rightarrow \mathcal{M}$ una aplicación descenso para una función objetivo suave f y suponer que para cada $x \in \mathcal{M}$, todos los puntos de acumulación de $\{F^{(k)}(x)\}_{k=1,2,\dots}$ son puntos críticos de f . Sea x_* un punto fijo de F (así x_* es un punto crítico de f). Suponer que x_* no es un mínimo local de f ; además suponer que existe una vecindad compacta \mathcal{U} de x_* donde, para cada punto crítico y de F en \mathcal{U} , se tiene que $f(y) = f(x_*)$. Entonces x_* es un punto inestable para F .*

Demostración.

Como x_* no es un mínimo local de f , entonces para cada vecindad \mathcal{V} de x_* , existe un punto $y \in \mathcal{V}$, tal que $f(y) < f(x_*)$. Ahora consideremos la sucesión $y_k := F^k(y)$. Luego con el propósito de establecer una contradicción, supongamos que y_k es estable, es decir, $y_k \in \mathcal{U}$ para todo k . Entonces, por compacidad $\{y_k\}$ tiene un punto de acumulación $z \in \mathcal{U}$. Por suposición, z es un punto crítico de f , por tanto $f(z) = f(x_*)$. Además, como F es una aplicación descenso, se tiene que

$$f(F^k(y)) \leq f(y),$$

pero como, $F^k(y) = y_k \rightarrow z$, entonces, $f(z) \leq f(y) < f(x_*)$, lo cual es una contradicción. ■

Ahora daremos un resultado de estabilidad.

Teorema 2.3 (Teorema de captura). *Sea $F : \mathcal{M} \rightarrow \mathcal{M}$ una aplicación descenso para una función objetivo suave f y suponer que, para cada $x \in \mathcal{M}$, todos los puntos de acumulación de $\{F^k(x)\}_{k=1,2,\dots}$ son puntos críticos de f . Sea x_* un minimizador local y un punto crítico aislado de f ; también suponer que $\text{dist}(F(x), x)$ tiende a cero conforme x tiende a x_* . Entonces x_* es un punto asintóticamente estable de F .*

Demostración. Sea \mathcal{U} una vecindad de x_* . Como x_* es un minimizador local aislado de f , existe una bola cerrada

$$\overline{B}_\varepsilon(x_*) := \{x \in \mathcal{M} : \text{dist}(x, x_*) \leq \varepsilon\},$$

tal que $\overline{B}_\varepsilon(x_*) \subset \mathcal{U}$ y $f(x) > f(x_*)$ para todo $x \in \overline{B}_\varepsilon(x_*) - \{x_*\}$. Por la condición sobre $\text{dist}(F(x), x)$, existe $\delta > 0$ tal que, para todo $x \in B_\delta(x_*)$, $F(x) \in \overline{B}_\varepsilon(x_*)$. Sea α el mínimo de f sobre el conjunto compacto $\overline{B}_\varepsilon(x_*) - B_\delta(x_*)$. Sea

$$\mathcal{V} = \{x \in \overline{B}_\varepsilon(x_*) : f(x) < \alpha\}.$$

Este conjunto está incluido en $B_\delta(x_*)$. Por tanto, para cada $x \in \mathcal{V}$, se cumple que $F(x) \in \overline{B}_\varepsilon(x_*)$, y también se cumple que $f(F(x)) \leq f(x) < \alpha$ puesto que F es una aplicación descenso. Luego se tiene que $F(x) \in \mathcal{V}$ para todo $x \in \mathcal{V}$, por tanto $F^{(n)}(x) \in \mathcal{V} \subset \mathcal{U}$ para todo $x \in \mathcal{V}$ y para todo n . Esto muestra la estabilidad. Además, como por suposición x_* es el único punto crítico de f en \mathcal{V} , se tiene que $\lim_{n \rightarrow \infty} F^{(n)}(x) = x_*$ para todo $x \in \mathcal{V}$, lo cual muestra la estabilidad asintótica. \blacksquare

2.5. Rapidez de convergencia

Hemos visto que, bajo ciertas suposiciones adecuadas, si la primera iteración del algoritmo 1 está suficientemente cerca a un minimizador local aislado x_* de f , entonces las sucesiones generadas $\{x_k\}$ converge a x_* . En esta sección estudiaremos cuán rápido dicha sucesión converge a x_* .

2.5.1. Orden de convergencia

Una sucesión $\{x_k\}_{k=0,1,\dots}$ de puntos de \mathbb{R}^n se dice que converge linealmente a un punto x_* si existe una constante $c \in (0, 1)$ y un entero $K \geq 0$ tal que, para todo $k \geq K$, se cumple $\|x_{k+1} - x_*\| \leq c \|x_k - x_*\|$. Para generalizar esta noción a variedades se puede hacer usando cartas y afirmamos que una sucesión $\{x_k\}_{k=0,1,\dots}$ de puntos de una variedad \mathcal{M} converge linealmente a un punto $x_* \in \mathcal{M}$ si, dado una carta (\mathcal{U}, ψ) con $x \in \mathcal{U}$, la sucesión $\{\psi(x_k)\}_{k=0,1,\dots}$ converge linealmente a $\psi(x_*)$; desafortunadamente esta noción no es independiente de la carta usada. (Para ver un ejemplo de la no independencia de la carta, ver P.-A. ABSIL, R. MAHONY, R. SEPULCHRE (2008), [1], pag. 69.) Sin embargo si \mathcal{M} es una variedad riemanniana, entonces la distancia riemanniana inducida nos hará posible definir la convergencia lineal de la siguiente manera.

Definición 2.6 (Convergencia lineal). Sea \mathcal{M} una variedad riemanniana y denotemos como dist la distancia riemanniana sobre \mathcal{M} . Decimos que una sucesión $\{x_k\}_{k=0,1,\dots}$ converge linealmente a un punto $x_* \in \mathcal{M}$ si existe una constante $c \in (0, 1)$ y un entero $K \geq 0$ tal que, para todo $k \geq K$, se cumple que

$$\text{dist}(x_{k+1}, x_*) \leq c \text{dist}(x_k, x_*). \quad (2.27)$$

El límite

$$\limsup_{k \rightarrow \infty} \frac{\text{dist}(x_{k+1}, x_*)}{\text{dist}(x_k, x_*)}$$

es llamado el factor de convergencia lineal de la sucesión. Un algoritmo iterativo sobre \mathcal{M} se dice que converge linealmente localmente a un punto x_* si existe una vecindad \mathcal{U} de x_* y una constante $c \in (0, 1)$ tal que, para cada punto inicial $x_0 \in \mathcal{U}$, la sucesión $\{x_k\}$ generada por el algoritmo satisface (2.27).

Nota 2.4. Una sucesión convergente $\{x_k\}$ sobre una variedad riemanniana \mathcal{M} converge linealmente a x_* con constante c si y solo si

$$\|R_{x_*}^{-1}(x_{k+1}) - R_{x_*}^{-1}(x_*)\| \leq c \|R_{x_*}^{-1}(x_k) - R_{x_*}^{-1}(x_*)\|$$

para todo k suficientemente grande, donde R es cualquier retracción sobre \mathcal{M} y $\|\cdot\|$ denota la norma sobre $T_{x_*}\mathcal{M}$ definida por la métrica riemanniana.

Definición 2.7. Sea \mathcal{M} una variedad y sea $\{x_k\}_{k=0,1,\dots}$ una sucesión sobre \mathcal{M} que converge a x_* . Sea (\mathcal{U}, ψ) una carta de \mathcal{M} con $x \in \mathcal{U}$. Si

$$\lim_{k \rightarrow \infty} \frac{\|\psi(x_{k+1}) - \psi(x_*)\|}{\|\psi(x_k) - \psi(x_*)\|} = 0,$$

entonces $\{x_k\}$ se dice que converge superlinealmente a x_* . Si existen constantes $p > 0$, $c \geq 0$, y $k \geq 0$ tal que, para todo $k \geq K$, se cumple que

$$\|\psi(x_{k+1}) - \psi(x_*)\| \leq c \|\psi(x_k) - \psi(x_*)\|^p, \quad (2.28)$$

entonces $\{x_k\}$ se dice que converge a x_* con orden al menos p . Un algoritmo iterativo sobre una variedad \mathcal{M} se dice que converge localmente a un punto x_* con orden al menos p si existe una carta (\mathcal{U}, ψ) en x_* y una constante $c > 0$ tal que, para cada punto inicial $x_0 \in \mathcal{U}$, la sucesión $\{x_k\}$ generada por el algoritmo satisface (2.28). Si $p = 2$, se dice que la convergencia es cuadrática, y cúbica si $p = 3$.

2.6. Tasa de convergencia de los métodos de búsqueda lineal

A continuación daremos una convergencia asintótica para el algoritmo 1 cuando η_k es elegido como $-\text{grad } f(x_k)$, sin ninguna otra suposición sobre como x_{k+1} es elegido.

El resultado involucra al menor y al mayor de los autovalores del Hessiano de f en un punto crítico x_* . Para poder hablar de los autovalores del Hessiano en un punto crítico daremos los siguientes resultados.

Lema 2.1. *Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ y $x_* \in \mathbb{R}^n$ tal que $Df(x_*) = 0$. Sea $F : \mathbb{R}^n \rightarrow \mathbb{R}^n$ y $y_* \in \mathbb{R}^n$ tal que $F(y_*) = x_*$ y que la matriz Jacobiana de F en y_* ,*

$$J_F(y_*) := \begin{bmatrix} \partial_1 F^1(y_*) & \cdots & \partial_n F^1(y_*) \\ \vdots & \ddots & \vdots \\ \partial_1 F^n(y_*) & \cdots & \partial_n F^n(y_*) \end{bmatrix},$$

es ortogonal (es decir, $J_F^T(y_)J_F(y_*) = I$). Sea H la matriz Hessiana de f en x_* ; es decir, $H_{ij} = \partial_i \partial_j f(x_*)$. Sea \hat{H} la matriz Hessiana de $f \circ F$ en y_* . Entonces $\lambda(H) = \lambda(\hat{H})$; es decir, el espectro de H y de \hat{H} son los mismos.*

Prueba.

Como $\partial_i(f \circ F)(y) = \sum_k \partial_k f(F(y)) \partial_j F^k(y)$, tenemos

$$\begin{aligned} \hat{H}_{ij} &= \partial_i \partial_j (f \circ F)(y_*) \\ &= \sum_{k,\ell} \partial_\ell \partial_k f(F(y_*)) \partial_i F^\ell(y_*) \partial_j F^k(y_*) + \sum_k \partial_k f(F(y_*)) \partial_i \partial_j F^k(y_*). \end{aligned}$$

Como x_* es un punto crítico de f , se obtiene que $\partial_k f(F(y_*)) = 0$. Por tanto tenemos en notación matricial

$$\hat{H} = J_F^T(y_*) H J_F(y_*) = J_F^{-1}(y_*) H J_F(y_*).$$

Esto muestra que H y \hat{H} tienen el mismo espectro, porque están relacionados por una transformación de similitud. ▀

Corolario 2.2. Sea f una función objetivo sobre una variedad riemanniana (\mathcal{M}, g) y sea $x_* \in \mathcal{M}$ un punto crítico de f , es decir, $\text{grad } f(x_*) = 0$. Sea (\mathcal{U}, ψ) cualquier carta tal que $x_* \in \mathcal{U}$ y que la representación de g_{x_*} en la carta es la identidad, es decir, $g_{ij} = \delta_{ij}$ en x_* . Entonces el espectro de la matriz Hessiana de $f \circ \psi^{-1}$ en $\psi(x_*)$ no depende de la elección de ψ .

Teorema 2.4. Sea $\{x_k\}$ una sucesión infinita de iteraciones generadas por el algoritmo 1 con $\eta_k := -\text{grad } f(x_k)$, que converge a un punto x_* . (Por el Teorema 2.1, x_* es un punto crítico de f .) Sea $\lambda_{H,\text{mín}}$ y $\lambda_{H,\text{máx}}$ el menor y el mayor de los autovalores de la Hessiana de f en x_* . Supongamos que $\lambda_{H,\text{mín}} > 0$ (así x_* es un minimizador local de f). Entonces, dado r en el intervalo $(r_*, 1)$ con $r_* = 1 - \min(2\sigma\bar{\alpha}\lambda_{H,\text{mín}}, 4\sigma(1-\sigma)\beta\frac{\lambda_{H,\text{mín}}}{\lambda_{H,\text{máx}}})$, existe un entero $K \geq 0$ tal que

$$f(x_{k+1}) - f(x_*) \leq (r + (1-r)(1-c))(f(x_k) - f(x_*)) \quad (2.29)$$

para todo $k \geq K$, donde c es el parámetro en el algoritmo 1.

Demostración.

Sea (\mathcal{U}, ψ) una carta de una variedad \mathcal{M} con $x_* \in \mathcal{U}$. Usamos la notación $\zeta_x := -\text{grad } f(x)$. Las expresiones coordinadas son denotadas con un sombrero, por ejemplo, $\hat{x} := \psi(x)$, $\hat{\mathcal{U}} = \psi(\mathcal{U})$, $\hat{f}(\hat{x}) := f(x)$, $\hat{\zeta}_{\hat{x}} := D\psi(x)[\zeta_x]$, $\hat{R}_{\hat{x}}(\hat{\zeta}) := \psi(R_x(\zeta))$. También denotemos como $y_{\hat{x}}$ al gradiente Euclidiano de \hat{f} en \hat{x} , es decir, $y_{\hat{x}} := (\partial_1 \hat{f}(\hat{x}), \dots, \partial_d \hat{f}(\hat{x}))^T$. Denotemos como $G_{\hat{x}}$ a la representación matricial de la métrica riemanniana en las coordenadas, y denotemos como $H_{\hat{x}_*}$ a la matriz Hessiana de \hat{f} en \hat{x}_* . Sin pérdida de generalidad, asumimos que $\hat{x}_* = 0$ y que $G_{\hat{x}_*} = I$ es la identidad matricial. El mayor trabajo es obtener, en una iteración actual x , una cota superior adecuada sobre $f(R_x(t^A \zeta_x))$, donde t^A es el paso de Armijo (entonces $t^A \zeta_x$ es el punto de Armijo). La condición de Armijo es

$$\begin{aligned} f(R_x(t^A \zeta_x)) &\leq f(x) - \sigma \langle \zeta_x, t^A \zeta_x \rangle \\ &\leq f(x) - \sigma t^A \langle \zeta_x, \zeta_x \rangle. \end{aligned} \quad (2.30)$$

Primero damos una cota inferior sobre $\langle \zeta_x, \zeta_x \rangle$ en términos de $f(x)$. Recordemos de (2.18) que $\hat{\zeta}_{\hat{x}} = G_{\hat{x}}^{-1} y_{\hat{x}}$, del cual se obtiene que

$$\langle \zeta_x, \zeta_x \rangle = \hat{\zeta}_{\hat{x}}^T G_{\hat{x}} \hat{\zeta}_{\hat{x}} = y_{\hat{x}} G_{\hat{x}}^{-1} y_{\hat{x}} = \|y_{\hat{x}}\|^2 (1 + O(\hat{x})), \quad (2.31)$$

ya que hemos asumido que G_0 es la identidad. Se deduce de $y_{\hat{x}} = H_0 \hat{x} + O(\hat{x}^2)$ y $\hat{f}_{\hat{x}} = \hat{f}(0) + \frac{1}{2} \hat{x}^T H_0 \hat{x} + O(\hat{x}^3)$ que, dado $\varepsilon \in (0, \lambda_{H, \text{mín}})$,

$$\hat{f}(\hat{x}) - \hat{f}(0) = \frac{1}{2} y_{\hat{x}}^T H_0^{-1} y_{\hat{x}} + O(\hat{x}^3) \leq \frac{1}{2} \frac{1}{\lambda_{H, \text{mín}} - \varepsilon} \|y_{\hat{x}}\|^2, \quad (2.32)$$

se cumple para todo \hat{x} suficientemente cerca a 0. De (2.31) y (2.32), concluimos que, dado $\varepsilon \in (0, \lambda_{H, \text{mín}})$,

$$f(x) - f(x_*) \leq \frac{1}{2} \frac{1}{\lambda_{H, \text{mín}} - \varepsilon} \langle \zeta_x, \zeta_x \rangle, \quad (2.33)$$

lo cual es la cota inferior deseada sobre $\langle \zeta_x, \zeta_x \rangle$. Usando (2.33) en (2.30) se obtiene

$$f(R_x(t^A \zeta_x)) - f(x_*) \leq (1 - 2(\lambda_{H, \text{mín}} - \varepsilon) \sigma t^A) (f(x) - f(x_*)) \quad (2.34)$$

Ahora encontraremos una cota inferior sobre el paso de Armijo t^A . Usamos la notación

$$\gamma_{\hat{x}, u}(t) := \hat{f}(\hat{R}_{\hat{x}}(tu))$$

y

$$h_x(t) = f(R_x(-t \zeta_x)).$$

observar que $h_x(t) = \gamma_{\hat{x}, -\hat{\zeta}_{\hat{x}}}(t)$ y que $\dot{h}_x(0) = -\langle \zeta_x, \zeta_x \rangle = \dot{\gamma}_{\hat{x}, -\hat{\zeta}_{\hat{x}}}(0)$, de donde se deduce que la condición de Armijo (2.30) se interpreta como

$$h_x(t^A) \leq h_x(0) - \sigma t^A \dot{h}_x(0). \quad (2.35)$$

Queremos encontrar una cota inferior sobre t^A . De la expansión de Taylor de h_x con el residuo en forma de Lagrange (ver [1], Apéndice A.6), se deduce que las t 's en el cual el lado derecho y el lado izquierdo de (2.35) son iguales, satisface

$$t = \frac{-2(1 - \sigma) \dot{h}_x(0)}{\ddot{h}_x(\tau)},$$

donde $\tau \in (0, t)$. En vista de la definición del punto de Armijo, concluimos que

$$t^A \geq \min \left(\bar{\alpha}, \frac{-2\beta(1-\sigma)\dot{h}_x(0)}{\max_{\tau \in (0, \bar{\alpha})} \ddot{h}_x(\tau)} \right) \quad (2.36)$$

Sea $B_\delta := \{\hat{x} : \|\hat{x}\| < \delta\}$ y $M := \sup_{\hat{x} \in B_\delta, \|u\|=1, t \in (0, \bar{\alpha})} \ddot{\gamma}_{\hat{x}, u}(t)$.

Entonces, $\max_{\tau \in (0, \bar{\alpha})} \ddot{h}_x(\tau) \leq M \|\hat{\zeta}_{\hat{x}}\|^2$. Se puede notar también que $\ddot{\gamma}_{\hat{x}, u}(0) = u^T H_0 u \leq \lambda_{H, \max} \|u\|^2$, así que $M \rightarrow \lambda_{H, \max}$ cuando $\delta \rightarrow 0$.

Finalmente, note que

$$\dot{h}_x(0) = -\hat{\zeta}_{\hat{x}}^T G_{\hat{x}} \hat{\zeta}_{\hat{x}} = \|\hat{\zeta}_{\hat{x}}\|^2 (1 + O(\hat{x})).$$

Usando estos resultados en (2.36) se obtiene que, dado $\varepsilon > 0$,

$$t^A \geq \min \left(\bar{\alpha}, \frac{2\beta(1-\sigma)}{\lambda_{H, \max} + \varepsilon} \right) \quad (2.37)$$

se cumple para todo x suficientemente cerca a x_* .

Ahora podemos combinar (2.37) y (2.34) para obtener una cota superior adecuada sobre $f(R_x(t^A \zeta_x))$:

$$f(R_x(t^A \zeta_x)) - f(x_*) \leq c_1 (f(x) - f(x_*)) \quad (2.38)$$

con

$$c_1 = 1 - \sigma \min \left(\bar{\alpha}, \frac{2\beta(1-\sigma)}{\lambda_{H, \max} + \varepsilon} \right) 2(\lambda_{H, \min} - \varepsilon).$$

Finalmente, la cota (2.38) junto con la cota (2.24) impuestas sobre el valor de f en la siguiente iteración, se obtiene

$$\begin{aligned} f(x_{k+1}) - f(x_*) &= f(x_{k+1}) - f(x_k) + f(x_k) - f(x_*) \\ &\leq -c(f(x_k) - f(R_{x_k}(t_k^A \zeta_{x_k}))) + f(x_k) - f(x_*) \\ &= (1-c)(f(x_k) - f(x_*)) + c(f(R_{x_k}(t_k^A \zeta_{x_k})) - f(x_*)) \\ &\leq (1-c+cc_1)(f(x_k) - f(x_*)) \\ &= (c_1 + (1-c_1)(1-c))(f(x_k) - f(x_*)), \end{aligned}$$

donde $c \in (0, 1)$ es la constante en la cota (2.24). ■

3. Aplicaciones del método de máximo descenso usando retracciones en variedades riemannianas

3.1. Minimización del cociente de Rayleigh sobre la esfera unitaria

En esta sección, como aplicación del método estudiado, implementaremos el Algoritmo 1, para el cociente de Rayleigh sobre la esfera, dado por

$$\begin{aligned} f: S^{n-1} &\rightarrow \mathbb{R} \\ x &\mapsto f(x) = x^T Ax, \end{aligned} \tag{2.39}$$

donde la matriz A es simétrica ($A = A^T$) pero no necesariamente definida positiva. También se mostrarán los resultados numéricos obtenidos por el algoritmo implementado en MATLAB, pero para tal objetivo primero daremos algunos resultados importantes que nos servirán para comprobar la eficiencia del método. Denotaremos por λ_1 al menor autovalor de A y por v_1 a un autovector unitario asociado a λ_1 .

3.1.1. Función objetivo y cálculo del gradiente

Consideremos la función

$$\begin{aligned} \bar{f}: \mathbb{R}^n &\rightarrow \mathbb{R} \\ x &\mapsto \bar{f}(x) = x^T Ax \end{aligned}$$

cuya restricción a la esfera unitaria S^{n-1} es la función (2.39).

Vemos a S^{n-1} como una subvariedad riemanniana del espacio euclidiano \mathbb{R}^n denotado con la métrica riemanniana canónica

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

Dado $x \in S^{n-1}$, tenemos

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

para todo $\zeta \in T_x \mathbb{R} \simeq \mathbb{R}^n$, del cual se deduce usando la definición del gradiente (2.17) que

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

El espacio tangente a S^{n-1} , visto como un subespacio de $T_x \mathbb{R}^n \simeq \mathbb{R}^n$, es

$$T_x S^{n-1} = \{\xi \in \mathbb{R}^n : x^T \xi = 0\}.$$

El espacio normal es

$$(T_x S^{n-1})^\perp = \{x\alpha : \alpha \in \mathbb{R}\}$$

Las proyecciones ortogonales sobre el espacio tangente y normal son

$$P_x \xi = \xi - xx^T \xi, \quad P_x^\perp \xi = xx^T \xi.$$

De la identidad (2.20), relacionada al gradiente sobre una subvariedad se tiene que

$$\text{grad } f(x) = 2P_x(Ax) = 2(Ax - xx^T Ax) \quad (2.40)$$

Las fórmulas obtenidas se resumirán en la siguiente tabla.

	Variedad (S^{n-1})	Espacio encajante (\mathbb{R}^n)
función objetivo	$f(x) = x^T Ax, x \in S^{n-1}$	$\bar{f}(x) = x^T Ax, x \in \mathbb{R}^n$
métrica	métrica inducida	$\bar{g}(\xi, \zeta) = \xi^T \zeta$
espacio tangente	$\xi \in \mathbb{R}^n : x^T \xi = 0$	\mathbb{R}^n
espacio normal	$\xi \in \mathbb{R}^n : \xi = \alpha x$	\emptyset
proyección sobre el espacio tangente	$P_x \xi = (I - xx^T)\xi$	identidad
gradiente	$\text{grad } f(x) = P_x \text{grad } \bar{f}(x)$	$\text{grad } \bar{f}(x) = 2Ax$
retracción	$R_x(\xi) = \text{qf}(x + \xi)$	$R_x(\xi) = x + \xi$

Tabla 2.1: Cociente de Rayleigh sobre la esfera unitaria

3.1.2. Puntos críticos del cociente de Rayleigh

Para analizar el algoritmo basado en la función objetivo dada por el cociente de Rayleigh sobre la esfera, lo primero que haremos es caracterizar a los puntos críticos.

Proposición 3.1. *Sea $A = A^T$ una matriz simétrica de orden $n \times n$. Un vector unitario $x \in \mathbb{R}^n$ es un autovector de A si y solo si es un punto crítico del cociente de Rayleigh (2.39).*

Prueba. Si x es un autovector unitario de A (es decir, $\|x\| = 1$), entonces $Ax = \lambda x$ para algún escalar λ , luego multiplicando por x^T a la izquierda se obtiene

$$x^T Ax = x^T \lambda x \Rightarrow \lambda \|x\|^2 = x^T Ax \Rightarrow \lambda = x^T Ax,$$

así se tiene

$$\lambda x = (x^T Ax)x \Rightarrow Ax = x(x^T Ax) \Rightarrow Ax - x(x^T Ax) = 0,$$

luego de (2.40), $\text{grad } f(x) = 0$. Por tanto x es un punto crítico del cociente de Rayleigh (2.39).

Recíprocamente, sea x un punto crítico de (2.39), es decir, $\text{grad } f(x) = 0$ con $x \in S^{n-1}$. Luego de la expresión de $\text{grad } f(x)$ (2.40), se deduce que x debe satisfacer

$$Ax = x(x^T Ax) \Rightarrow Ax = (x^T Ax)x,$$

como $\lambda = x^T Ax \in \mathbb{R}$, entonces λ es un escalar que cumple $Ax = \lambda x$. Por tanto x es un autovector de A . ■

Proposición 3.2. *Sea $A = A^T$ una matriz simétrica con autovalores $\lambda_1 \leq \dots \leq \lambda_n$ y los autovectores ortonormales v_1, \dots, v_n asociados a $\lambda_1, \dots, \lambda_n$ respectivamente. Entonces*

- (i) $\pm v_1$ son minimizadores locales y globales del cociente de Rayleigh (2.39); si el autovalor λ_1 es simple, entonces $\pm v_1$ son solo minimizadores.
- (ii) $\pm v_n$ son maximizadores locales y globales de (2.39); si el autovalor λ_n es simple, entonces $\pm v_n$ son solo maximizadores.

(III) $\pm v_q$ correspondientes a los autovalores interiores (i.e., $\lambda_2, \dots, \lambda_{n-1}$) son puntos silla de (2.39).

Prueba. El ítem (I) se deduce de la proposición 2.1.1 de [1]. El ítem (II) se deduce de la misma proposición por reemplazar A por $-A$ y se nota que cambia máximo con mínimo y los autovectores de más a la izquierda con los autovectores de más a la derecha. Para el ítem (III), sea v_q un autovector correspondiente a un autovalor interior λ_q y considerar la curva

$$\begin{aligned} \gamma: \mathbb{R} &\rightarrow \mathbb{R}^n \\ t &\mapsto \gamma(t) = \frac{v_q + tv_1}{\|v_q + tv_1\|}, \end{aligned}$$

por simples cálculos se puede mostrar que

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

Así mismo, para la curva

$$\begin{aligned} \gamma: \mathbb{R} &\rightarrow \mathbb{R}^n \\ t &\mapsto \gamma(t) = \frac{v_q + tv_n}{\|v_q + tv_n\|} \end{aligned}$$

tenemos

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

De esto se deduce que v_q es un punto silla del cociente de Rayleigh f . ■

De la Proposición 3.1 y del análisis de convergencia de los métodos de búsqueda lineal (Teorema 2.1) se deduce que todos los métodos dentro de la clase del Algoritmo 1 producen iteraciones que converge al conjunto de autovectores de A . Además por la Proposición 3.1 y ya que estamos considerando métodos de descenso, se deduce que si λ_1 es simple, la convergencia es estable a $\pm v_1$ e inestable a todos los otros autovectores.

De aquí en adelante consideraremos para los ejemplos del Algoritmo 1 a η_k como

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

Es claro que esta elección de la dirección de búsqueda es el gradiente relacionado. Luego tomaremos una retracción; una elección razonable es

$$R_x(\xi) = \frac{x + \xi}{\|x + \xi\|}, \quad (2.41)$$

donde $\|\cdot\|$ denota la norma euclidiana en \mathbb{R}^n , además $\|y\| := \sqrt{y^T y}$.

3.1.3. Búsqueda lineal de Armijo

Ya visto los resultados anteriores, ahora ya tenemos todo lo necesario para aplicar el método, dado en el Algoritmo 1, para resolver el problema de minimizar el cociente de Rayleigh sobre la esfera S^{n-1} . El algoritmo que se obtiene para este ejemplo es el siguiente.

Algoritmo 2 Búsqueda lineal de Armijo para el cociente de Rayleigh sobre S^{n-1}

Entrada: Una matriz simétrica A , escalares $\bar{\alpha} > 0$, β , $\sigma \in (0, 1)$ y un punto inicial

$$x_0 \in \mathbb{R}^n \text{ tal que } \|x_0\| = 1.$$

Salida: Sucesión de iteraciones $\{x_k\}$.

- 1: **para** $k = 0, 1, 2, \dots$ **hacer**
- 2: Calcular $\eta_k = -2(Ax_k - x_k x_k^T A x_k)$.
- 3: Encontrar el menor entero $m \geq 0$ tal que

$$f(R_{x_k}(\bar{\alpha}\beta^m \eta_k)) \leq f(x_k) - \sigma \bar{\alpha} \beta^m \eta_k^T \eta_k,$$

con f definida en (2.39) y R definida en (2.41).

- 4: Usar la fórmula de actualización

$$x_{k+1} = R_{x_k}(\bar{\alpha}\beta^m \eta_k).$$

- 5: **fin de para**
-

3.1.4. Implementación del Algoritmo 2 en MATLAB

A continuación se presentará el algoritmo en código Matlab que resuelve el problema de minimizar el cociente de Rayleigh sobre la esfera unitaria.

```

fprintf('\t\t Algoritmo de máximo descenso usando
        retracciones en variedades riemannianas\n');
fprintf('\t\t =====\n\n');
Y=input('\n Ingrese un vector Y: ');
%debe ser un vector columna de i componentes
fprintf('\n la norma N de Y es: ')
N=norm(Y)
fprintf('\n El vector unitario X de Y es: ');
X=Y/norm(Y)
A=input('\n Ingrese una matriz simétrica A: ')
%debe ser una matriz de orden i*i
alpha=1;
sigma=0.1;
beta=0.5;
epsilon=0.00001; % error de aproximación permitido
Nmax=100;
% esto es el número máximo de iteraciones que se realizarán
k=1;
testArmijo=0;
fprintf('\n El punto inicial x(i) es: %d',X)
%i varia de 1 hasta el número de componentes que tenga X
while k<=Nmax,
fprintf('\n\n =====');
fprintf('\n Iteración N°: %d',k)
fprintf('\n =====\n');
fprintf('\n El punto x(i) para esta iteración es : %d',X)
gradf=2*(A*X-X*X'*A*X)
Normadf=norm(gradf)
    if Normadf >= epsilon
        m=0;
        factual=X'*A*X

```

```

eta=-2*(A*X-X*X'*A*X); % dirección de búsqueda
R=(X+eta)/norm(X+eta); % esta es la retracción tomada
fnuevo=R'*A*R ;
dif=factual-fnuevo
t=1;
test=sigma*t*eta'*eta
testArmijo=testArmijo+1;
while dif < test ,
    m = m+1
    testArmijo=testArmijo+1;
    t=beta^m*alpha; % tamaño de paso de Armijo
    r=(X+t*eta)/norm(X+t*eta)
    % la retracción se va actualizando en cada iteración
    dif=factual-r'*A*r
    test=sigma*t*eta'*eta
end
X=(X+t*eta)/norm(X+t*eta)
% fórmula de actualización para cada iteración
fprintf('\n Nuevo punto x(i) es=: %d',X)
k=k+1;
else
    fprintf('\n criterio de error satisfecho Normagradiante=:
           %d',Normadf)
    fprintf('\n el punto óptimo x(i) aproximado es:% d',X)
    fprintf('\n el valor óptimo aproximado es:% d',factual)
    fprintf('\n Número de iteraciones es:% d',k-1)
    fprintf('\n Número de test de Armijo:% d',testArmijo)
    k=1000000;
end
end
if k >999999

```

```

        fprintf('\n Resolvimos el problema: %d')
else
    fprintf('\n : Número de iteraciones excedido a %d',k-1)
    fprintf('\n Número de test de Armijo: %d',testArmijo)
end

```

A continuación se presentarán los resultados computacionales de aplicar el algoritmo implementado para resolver el problema de minimizar (2.39) para el caso particular

$$A = \begin{bmatrix} 2 & 5 \\ 5 & 1 \end{bmatrix},$$

es decir

$$\min f(x) = x^T \begin{bmatrix} 2 & 5 \\ 5 & 1 \end{bmatrix} x$$

Para los datos de entrada necesarios en el algoritmo 2 son considerados $\sigma = 0.1$, $\beta = 0.5$, $\bar{\alpha} = 1$ y como punto inicial tomaremos a $\begin{bmatrix} 0.6 & 0.8 \end{bmatrix}^T$ que es el vector unitario obtenido de $\begin{bmatrix} 3 & 4 \end{bmatrix}^T$. Con estos datos los resultados computacionales son los presentados en la siguiente tabla.

x_k	Iteración	Punto Óptimo	Valor Óptimo	$\ \text{grad } f(x_k)\ $
x_0	1	(0.600000, 0.800000)	6.160000	3.760000
x_1	2	(-0.618911, 0.785461)	-3.478254	1.366731
x_2	3	(-0.683516, 0.729935)	-3.522032	0.341732
x_3	4	(-0.667774, 0.744364)	-3.524748	0.087431
x_4	5	(-0.671831, 0.740704)	-3.524925	0.022401
x_5	6	(-0.670794, 0.741644)	-3.524937	0.005740
x_6	7	(-0.671060, 0.741403)	-3.524938	0.001471
x_7	8	(-0.670991, 0.741465)	-3.524938	3.7685e-004
x_8	9	(-0.671009, 0.741449)	-3.524938	9.6562e-005
x_9	10	(-0.671004, 0.741453)	-3.524938	2.4743e-005
x_{10}	11	(-0.671006, 0.741452)	-3.524938	6.3399e-006

Tabla 2.2: Resultados numéricos de minimizar el cociente de Rayleigh

Para ver que estos resultados estan de acuerdo con los resultados teóricos dados anteriormente primeramente determinemos los autovalores de A

$$\det(A - \lambda I) = \det \begin{bmatrix} 2 - \lambda & 5 \\ 5 & 1 - \lambda \end{bmatrix} = (2 - \lambda)(1 - \lambda) - 25 = 0$$

$$\Rightarrow \lambda^2 - 3\lambda - 23 = 0$$

Entonces los autovalores son: $\lambda_1 = -3.5249$ y $\lambda_2 = 6.5249$.

Luego para λ_1 tenemos que

$$A - \lambda_1 I = \begin{bmatrix} 5.5249 & 5 \\ 5 & 4.5249 \end{bmatrix}.$$

Ahora si

$$(v_1, v_2) \in \text{Nuc}(A - \lambda_1 I) \Rightarrow \begin{bmatrix} 5.5249 & 5 \\ 5 & 4.5249 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Rightarrow \begin{cases} 5.5249v_1 + 5v_2 = 0 \\ 5v_1 + 4.5249v_2 = 0 \end{cases}$$

$$\Rightarrow 0.5249v_1 + 0.4751v_2 = 0,$$

entonces el autoespacio de A asociado a λ_1 es

$$\text{Nuc}(A - \lambda_1 I) = \{(v_1, v_2) \in \mathbb{R}^2 : 0.5249v_1 + 0.4751v_2 = 0\}.$$

Luego por la Proposición 3.2 se debe cumplir que el punto óptimo obtenido $x^* = (-0.671006, 0.741452)$ debe pertenecer al autoespacio de A asociado a λ_1 , veamos:

$$0.5249(-0.671006) + 0.4751(0.741452) = 0.000053 \approx 0,$$

vemos que la ecuación del autoespacio al ser evaluada en el punto óptimo obtenido por el Matlab se aproxima a cero; esto sucede ya que dicho punto es solo una aproximación al óptimo real, puesto que hemos considerado un error de aproximación de $\varepsilon = 0.00001$.

De esta manera podemos concluir que el punto $x^* = (-0.671006, 0.741452)$ es un autovector asociado a λ_1 y es el óptimo para el ejemplo considerado, verificando así la Proposición 3.2.

3.2. Minimización de la función de Brockett sobre la variedad de Stiefel

Ahora minimizaremos una función objetivo definida como una suma ponderada $\sum_i \mu_i x_{(i)}^T A x_{(i)}$ de cocientes de Rayleigh sobre la esfera bajo la restricción de ortogonalidad $x_{(i)}^T x_{(i)} = \delta_{ij}$.

3.2.1. Función objetivo y dirección de Búsqueda

La función objetivo admite la siguiente expresión en forma matricial:

$$\begin{aligned} f : \text{St}(p, n) &\rightarrow \mathbb{R} \\ X &\mapsto \text{tr}(X^T A X N), \end{aligned} \quad (2.42)$$

donde $N = \text{diag}(\mu_1, \dots, \mu_p)$, con $0 \leq \mu_1 \leq \dots \leq \mu_p$, y $\text{St}(p, n)$ denota la variedad de Stiefel ortogonal

$$\text{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\}.$$

Como en la sección 1.5.1, vimos a $\text{St}(p, n)$ como una subvariedad regular del espacio euclidiano $\mathbb{R}^{n \times p}$. El espacio tangente es (ver ejemplo 1.5)

$$T_X \text{St}(p, n) = \{Z \in \mathbb{R}^{n \times p} : X^T Z + Z^T X = 0\}$$

Además se considera a $\text{St}(p, n)$ como una subvariedad riemanniana de $\mathbb{R}^{n \times p}$ dotada con el producto interno canónico

$$\langle Z_1, Z_2 \rangle := \text{tr}(Z_1^T Z_2).$$

También el espacio normal a $\text{St}(p, n)$ en un punto X es

$$(T_x \text{St}(p, n))^\perp = \{XS : S^T = S\}.$$

La proyección ortogonal P_x sobre $T_X \text{St}(p, n)$ es dado por

$$P_X Z = Z - X \text{sim}(X^T Z) = (I - X X^T) Z + X \text{asim}(X^T Z),$$

donde $\text{sim}(M) := \frac{1}{2}(M + M^T)$ y $\text{asim}(M) = \frac{1}{2}(M - M^T)$ denotan la parte simétrica y la parte antisimétrica de la descomposición de M en un término simétrico y un antisimétrico.

Considerar la función

$$\begin{aligned}\bar{f} : \mathbb{R}^{n \times p} &\rightarrow \mathbb{R} \\ X &\mapsto \text{tr}(X^T A X N),\end{aligned}$$

así se tiene que $f = \bar{f}|_{\text{St}(p,n)}$. Luego tenemos

$$D\bar{f}(X)[Z] = 2 \text{tr}(Z^T A X N),$$

así

$$\text{grad } \bar{f}(X) = 2 A X N$$

y

$$\begin{aligned}\text{grad } f(X) &= P_X \text{grad } \bar{f}(X) \\ &= 2 A X N - 2 X \text{sim}(X^T A X N) \\ &= 2 A X N - X X^T A X N - X N X^T A X.\end{aligned}$$

Ahora nos falta elegir una retracción, para lo cual elegiremos la retracción propuesta en el ejemplo 2.2, es decir,

$$R_X(\xi) = \text{qf}(X + \xi).$$

Todas las fórmulas dadas se resumirán en la siguiente tabla.

	Variedad (S^{n-1})	Espacio encajante (\mathbb{R}^n)
función objetivo	$\text{tr}(X^T A X N)$, $X^T X = I_p$	$\text{tr}(X^T A X N)$, $X \in \mathbb{R}^{n \times p}$
métrica	métrica inducida	$\langle Z_1, Z_2 \rangle = \text{tr}(Z_1^T Z_2)$
espacio tangente	$Z \in \mathbb{R}^{n \times p} : \text{sim}(X^T Z) = 0$	$\mathbb{R}^{n \times p}$
espacio normal	$Z \in \mathbb{R}^{n \times p} : Z = X S, S^T = S$	\emptyset
proyección sobre el espacio tangente	$P_X Z = Z - X \text{sim}(X^T Z)$	identidad
gradiente	$\text{grad } f(x) = P_x \text{grad } \bar{f}(X)$	$\text{grad } \bar{f}(x) = 2 A X N$
retracción	$R_X(\xi) = \text{qf}(X + Z)$	$R_X(Z) = X + Z$

Tabla 2.3: Función de Brockett sobre la variedad de Stiefel

3.2.2. Implementación del algoritmo que minimiza la función de Brockett sobre la variedad de Stiefel

```
format long
fprintf('\t\t Algoritmo de máximo descenso usando
        retracciones en variedades riemannianas\n');
fprintf('\t\t =====\n\n');
fprintf('\t\t Minimización de la función de Brockett
        sobre la variedad de Stiefel\n');
fprintf('\t\t =====\n\n');
M=input('\n Ingrese una matriz M: ');
%debe ser una matriz de orden n*p
[Q,R]=qr(M)% es la descomposición QR de M
A=input('\n Ingrese una matriz simétrica A: ');
%debe ser una matriz de orden n*n
X = Q;
%es la matriz ortogonal que se usará como punto inicial
u=input('\n Ingrese un vector u tal que 0<= u_1<=...<=u_p: ');
%debe ser un vector de orden 1*p tal que 0<= u_1<=...<=u_p
N=diag(u)
alpha=1;
sigma=0.1;
beta=0.5;
epsilon=0.00001; %indica el error de aproximación permitido
Nmax=100;
%indica el número máximo de iteraciones que se realizarán
k=1;
testArmijo=0;
fprintf('\n El punto inicial x(i) es: %d',X)
%i varia de 1 hasta el N° de entradas que tenga la matriz X
while k<=Nmax,
    fprintf('\n\n =====');
```

```

fprintf('\n Iteración No: %d',k)
fprintf('\n =====\n');
fprintf('\n El punto X(i) para esta iteración es : %d',X)
gradf=2*(A*X*N)-(X*X'*A*X*N)-(X*N*X'*A*X)
Normadf=norm(gradf)
    if Normadf >= epsilon
m=0;
factual=trace(X'*A*X*N)
eta=-2*(A*X*N)-(X*X'*A*X*N)-(X*N*X'*A*X);
% esta es la dirección de búsqueda
[S,T]=qr(X+eta);
Re=S % esta es la retracción tomada
fnuevo=trace(Re'*A*Re*N)
dif=factual-fnuevo
t=1;
test=sigma*t*eta'*eta
testArmijo=testArmijo+1;
    while dif < test ,
        m = m+1
        testArmijo=testArmijo+1;
        t=beta^m*alpha;
        % este viene a ser el tamaño de paso de Armijo
        [W,Z]=qr(X+t*eta)
        r = W
        % la retracción se va actualizando en cada iteración
        dif=factual-trace(r'*A*r*N)
        test=sigma*t*eta'*eta
    end
[W,Z]=qr(X+t*eta);
X = W
% esta es la fórmula de actualización para cada iteración

```

```

    fprintf('\n Nuevo punto x(i) es=: %d',X)
    k=k+1;
else
    fprintf('\n criterio de error satisfecho Normagradiante=:
           %d',Normadf)
    fprintf('\n el punto óptimo x(i) aproximado es: %d',X)
    fprintf('\n el valor óptimo aproximado es:% d',factual)
    fprintf('\n Número de iteraciones es: %d',k-1)
    fprintf('\n Número de test de Armijo: %d',testArmijo)
    fprintf('\n el punto óptimo es: ')
    k=1000000;
end
end
if k >999999
    fprintf('\n Resolvimos el problema: %d')
else
    fprintf('\n : Número de iteraciones excedido a %d',k-1)
    fprintf('\n Número de test de Armijo: %d',testArmijo)
end
end

```

A continuación se presentarán los resultados computacionales que se obtienen al aplicar el algoritmo implementado para resolver el problema de minimizar

$$\begin{aligned}
 f: \text{St}(p, n) &\rightarrow \mathbb{R} \\
 X &\mapsto \text{tr}(X^T A X N),
 \end{aligned}$$

para

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 5 \end{bmatrix} \text{ y } N = \begin{bmatrix} 4 & 0 \\ 0 & 5 \end{bmatrix}$$

x_k	Nº iter.	Punto Óptimo	Valor Ópt.	$\ \text{grad } f(x_k)\ $
x_0	1	$\begin{bmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{bmatrix}$	31.920000	1.440000
x_1	2	$\begin{bmatrix} -4.421817e-01 & -8.969255e-01 \\ -8.969255e-01 & 4.421817e-01 \end{bmatrix}$	30.586574	1.189812
x_2	3	$\begin{bmatrix} 1.743662e-01 & 9.846809e-01 \\ 9.846809e-01 & -1.743662e-01 \end{bmatrix}$	30.091211	0.515085
x_3	4	$\begin{bmatrix} -6.276994e-02 & -9.980280e-01 \\ -9.980280e-01 & 6.276994e-02 \end{bmatrix}$	30.011820	0.187938
x_4	5	$\begin{bmatrix} 2.229207e-02 & 9.997515e-01 \\ 9.997515e-01 & -2.229207e-02 \end{bmatrix}$	30.001491	0.066860
x_5	6	$\begin{bmatrix} -7.902974e-03 & -9.999688e-01 \\ -9.999688e-01 & 7.902974e-03 \end{bmatrix}$	30.000187	0.023708
x_6	7	$\begin{bmatrix} 2.801143e-03 & 9.999961e-01 \\ 9.999961e-01 & -2.801143e-03 \end{bmatrix}$	30.000024	0.008403
x_7	8	$\begin{bmatrix} -9.928140e-04 & -9.999995e-01 \\ -9.999995e-01 & 9.928140e-04 \end{bmatrix}$	30.000003	0.002978
x_8	9	$\begin{bmatrix} 3.518836e-04 & 9.999999e-01 \\ 9.999999e-01 & -3.518836e-04 \end{bmatrix}$	30.000000	0.001056
x_9	10	$\begin{bmatrix} -1.247183e-04 & -1.000000e+00 \\ -1.000000e+00 & 1.247183e-04 \end{bmatrix}$	30.000000	3.741548e-004
x_{10}	11	$\begin{bmatrix} 4.420394e-05 & 1.000000e+00 \\ 1.000000e+00 & -4.420394e-05 \end{bmatrix}$	30.000000	1.326118e-004
x_{11}	12	$\begin{bmatrix} -1.566722e-05 & -1.000000e+00 \\ -1.000000e+00 & 1.566722e-05 \end{bmatrix}$	30.000000	4.700166e-005
x_{12}	13	$\begin{bmatrix} 5.552938e-06 & 1.000000e+00 \\ 1.000000e+00 & -5.552938e-06 \end{bmatrix}$	30.000000	1.665881e-005

Tabla 2.4: Resultados numéricos de minimizar la función de Brockett

CAPÍTULO III

VARIABLES E HIPÓTESIS

1. Variable de la investigación

$x_{k+1} = R_{x_k}(t_k \eta_k)$; donde $\eta_k \in T_{x_k} \in T_{x_k} \mathcal{M}$, t_k es un escalar y R es la aplicación Retracción.

2. Operacionalización de variables

Variable	Definición conceptual	Definición operacional	Dimensiones	Indicadores
$x_{k+1} = R_{x_k}(t_k \eta_k)$	Sucesión de puntos x_k que genera el método de máximo descenso en variedades riemannianas usando retracciones.	Fórmula de iteración, la cual requiere de un punto inicial para generar todas las demás iteraciones.	- Implementar el algoritmo	- Interpretar el algoritmo propuesto. - Programar el algoritmo en MATLAB. - Determinar todos los parámetros requeridos en la fórmula de iteración.
			- Ejecutar el algoritmo	- Escoger un punto inicial x_0 . - Dar un error de aproximación. - Determinar el óptimo.

3. Hipótesis general e hipótesis específica

Hipótesis general

Conociendo los métodos clásicos de búsqueda lineal generalizado a variedades riemannianas, en específico el método de máximo descenso usando retracciones de los cuales conocemos su convergencia y velocidad de convergencia, el algoritmo de dicho método se puede implementar en el software MATLAB de manera que se obtenga resultados muy cercanos al valor óptimo, y así se pueda aplicar para resolver problemas de optimización de la vida real usando dicho software.

Hipótesis específica

Para implementar el algoritmo de nuestro método en estudio es necesario saber determinar la retracción de la variedad sobre la cual trabajamos, por tanto nosotros nos enfocamos en variedades matriciales de las cuales hallamos su retracción.

CAPÍTULO IV

METODOLOGÍA

1. Tipo de investigación

El estudio de la investigación es de carácter teórico-aplicativo. En el desarrollo del trabajo si se obtiene los resultados deseados, esto será un nuevo aporte científico, muy significativo en optimización en la búsqueda de algoritmos más eficientes tanto teórica como computacionalmente, para la solución de problemas prácticos que surgen en diversas áreas de las ciencias e ingeniería, y en lo posible que sirva como una motivación para que se siga investigando en esta área de la matemática y se desarrollen nuevos métodos basados en la búsqueda lineal.

2. Diseño de la investigación

Durante el desarrollo del proyecto utilizamos un método inductivo-deductivo, es decir, generalizamos a variedades las definiciones, teoremas, y corolarios de resultados clásicos de optimización y a la vez particularizamos los resultados obtenidos aplicándolos a algunos ejemplos específicos para los cuales implementamos su algoritmo.

Para tal objetivo en la primera parte presentamos la teoría básica de variedades diferenciables, de las cuales se estudia su topología, su diferenciabilidad entre otras propiedades y se da algunos ejemplos para un mejor entendimiento, entre ellos estudiamos la variedad de Stiefel y la esfera unitaria que los utilizaremos en las aplicaciones del método y por último se dió una introducción a las variedades riemannianas.

En un segundo paso, definimos la aplicacion **retracción**, el cuál es un paso clave para estudiar el método de máximo descenso, probamos que la sucesión generada por el método, usando la regla de búsqueda generalizada de Armijo, converge a un

punto crítico de la función objetivo y también se estudió la convergencia y velocidad de convergencia del método.

Por último, aplicamos el método de máximo descenso para minimizar el cociente de Rayleigh sobre la esfera unitaria y también para minimizar la función de Brocket sobre la variedad de Stiefel de los cuales implementamos el algoritmo y finalmente se presenta los resultados numéricos obtenidos en MATLAB.

3. Población y muestra

Por ser nuestro trabajo teórico-aplicativo, no se estudia ninguna población, sin embargo, nuestro trabajo se encuentra dentro del universo de los métodos de optimización en variedades riemannianas.

4. Técnicas e instrumentos de recolección de datos

Para la realización de nuestro trabajo de tesis se revisó bibliografía especializada y recopilación de información obtenida vía internet relacionada al tema de interés.

5. Procedimiento de recolección de datos

Por ser nuestro trabajo netamente abstracto, no se necesitó procedimientos de recolección de datos más que la revisión de bibliografía (libros, páginas web, papers, etc.)

6. Procesamiento estadístico y análisis de datos

Por la característica del trabajo no se realizó ningún análisis estadístico.

CAPÍTULO V

RESULTADOS

Los resultados que se obtuvieron en el siguiente trabajo son los siguientes:

- Introducimos la noción de una retracción R en un punto x denotado por R_x , definida como:

$$\begin{aligned} R_x : T_x \mathcal{M} &\rightarrow \mathcal{M} \\ \xi &\mapsto R_x(\xi), \end{aligned}$$

lo cual es un paso clave para estudiar el método de máximo descenso sobre variedades riemannianas usando dicha aplicación.

- Introducimos el método de máximo descenso usando retracciones en variedades riemannianas y probamos que todos los puntos de acumulación de la sucesión de iteraciones generada por el algoritmo dado es un punto crítico de la función objetivo, siendo este hecho de suma importancia para analizar la convergencia.
- También dimos un resultado de convergencia para el algoritmo dado en el trabajo, dicho resultado involucra al menor y al mayor autovalor del Hessiano en un punto crítico.
- Aplicamos el método desarrollado para minimizar el cociente de Rayleigh sobre la esfera unitaria y también para minimizar la función de Brocket sobre la variedad de Stiefel e implementamos sus respectivos algoritmos en MATLAB.

CAPÍTULO VI

DISCUSIÓN DE RESULTADOS

- Los resultados de convergencia, para el método de máximo descenso dado en este trabajo son una generalización del método clásico de máximo descenso en el espacio euclidiano y se obtienen gracias al análisis de estabilidad de puntos fijos, puesto que solo se usa información de primer orden sobre la función objetivo.
- Los aportes dados en la investigación, nos motivan a desarrollar nuevos métodos de optimización que usen una búsqueda lineal para determinar su tamaño de paso.
- Una interrogante que deja este trabajo es cuán eficiente es el método al ser aplicado a problemas de la vida cotidiana, es por eso que este trabajo debe servir como motivación a encontrar problemas reales en diversas áreas de ciencias en ingenierías, que sean desarrollados por el método estudiado.
- Los resultados numéricos obtenidos en MATLAB fueron verificados teóricamente y observamos que los resultados coinciden, por lo tanto, podemos decir que nuestro algoritmo está correctamente implementado.

CAPÍTULO VII

CONCLUSIONES

Al finalizar el trabajo se llegaron a las siguientes conclusiones:

- Si usamos el método de máximo descenso, desarrollado en este trabajo, para resolver problemas de optimización sobre variedades riemannianas, tendremos la seguridad que si cumple ciertas condiciones adecuadas en la primera iteración, estaremos lo suficientemente próximo al óptimo.
- Los métodos de búsqueda lineal en optimización son importantes para determinar la longitud de paso, en cada iteración, de alguna dirección seleccionada. Con respecto al método estudiado en este trabajo podemos decir que es de vital importancia una adecuada selección del tamaño de paso del método para así obtener algoritmos eficientes.
- Para un estudio más profundo de los algoritmos de búsqueda lineal sobre variedades riemannianas, entre ellos nuestro método estudiado, es necesario tener conocimientos de diferentes áreas de la matemática, como por ejemplo de la Geometría Riemanniana, Álgebra Matricial Computacional, Análisis en Espacios Euclidianos y Álgebra Lineal.
- Usando el método de máximo descenso con retracciones y una búsqueda lineal se obtuvo la convergencia del método hacia un punto crítico de la función objetivo, donde la función objetivo es una función diferenciable no necesariamente convexa.

CAPÍTULO VIII

RECOMENDACIONES

- Dentro de un proyecto de investigación como este, siempre se desea que haya una mejora continua del mismo, como por ejemplo hacer la implementación en otro lenguaje de programación como C++. También se puede implementar los mismos ejemplos de nuestro trabajo usando otro método en variedades riemannianas y comparar los resultados para determinar que método es más eficiente. Por lo tanto se recomienda a futuros estudiantes que tengan interés en investigar, desarrollen los puntos mencionados.
- Nuestro método de optimización estudiado no es el único método que utilizan una búsqueda lineal para llegar al óptimo, también existen otros métodos como el método de Newton, Método Cuasinewton y método de Regiones de Confianza; por lo tanto se recomienda a los lectores interesados en revisar ABSIL P.-A., MAHONY R. and SEPULCHRE R. (2008), [1], pues allí encontrarán una teoría amplia y muy detallada acerca de los métodos mencionados.

BIBLIOGRAFÍA

- [1] ABSIL P.-A., MAHONY R. and SEPULCHRE R. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, New Jersey, 2008.
- [2] ABSIL P.-A., MAHONY R. and SEPULCHRE R. *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*. Acta Applicandae Mathematicae, Volume 80, Issue 2, pp. 199-220, 2004.
- [3] ABSIL P.-A. and JÉRÔME MALICK. *Projection-like retractions on matrix manifolds*. Technical report UCL-INMA-2010.038, 2010.
- [4] ABRAHAM R., MARSDEN J.E. and RATIU T. *Manifolds, Tensor Analysis, and Applications*. Springer-Verlag Publishing Company, Inc, New York, third edition, 2001.
- [5] BARRET O'NEILL. *Semi-Riemannian Geometry*, volume 103 of Pure and Applied Mathematics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983.
- [6] BAZARAA M., SHERALY H. and SHETTY C. *Nonlinear Programming*. Jon Wiley & Sons, third edition, 2006.
- [7] BENTO G.C.¹, da CRUZ NETO J.X.² and OLIVEIRA P.R.³. *Convergence of inexact descent methods for nonconvex optimization on Riemannian manifolds*. ¹IME, Universidade Federal de Goiás, Goiânia, GO 74001-970, BR., ²Centro de Ciências da Natureza, DM, Universidade Federal do Piauí, Teresina, PI 64049-550, BR., ³COPPE-Sistemas, Universidade Federal do Rio de Janeiro, Rio de Janeiro, RJ 21945-970, BR, 2011.
- [8] BERTSEKAS D.P. *Nonlinear Programming*. Massachusetts institute of technology, second edition, 1999.
- [9] BOOTHBY W.M. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, London, 1986.

- [10] BROCKETT R.W. *Differential geometry and the design of gradient algorithms*. In *Differential geometry: partial differential equations on manifolds* (Los Angeles, CA, 1990), volume 54 of Proc. Sympos. Pure Math., pages 69-92. Amer. Math. Soc., Providence, RI, 1993.
- [11] do CARMO M.P. *Geometria Diferencial de Curvas e Superficies*, Rio de Janeiro: Sociedade Brasileira de Matemática, terceira edição,
- [12] do CARMO M.P. *Riemannian geometry*. Mathematics: Theory & Applications. Birkhäuser Boston Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.
- [13] GABAY D. *Minimizing a differentiable function over a differentiable manifold*. *Journal of Optimization Theory and Application*, 1982, vol 37, pp. 177-219.
- [14] LUENBERGER D.G. *Introduction to Linear and Nonlinear Programming*. Addison-Wesley, Reading, MA, 1973.
- [15] LUENBERGER D.G. *The gradient projection method along geodesics*. *Management Science*, 1972, Vol 18 n. 1, pp. 620-631.
- [16] MAHONY R. *Optimization Algorithms on Homogeneous Spaces: with Applications in Linear Systems Theory*. PhD thesis, Department of Systems Engineering, Australian National University, Cambridge, 1994.
- [17] MOORE J.B., MAHONY R.E. and HELMKE U. *Numerical gradient algorithms for eigenvalue and singular value calculations*. *SIAM J. Matrix Anal. Appl.*, 15(3):881-902, 1994.
- [18] PAPA QUIROZ E.A., QUISPE E.M. and OLIVEIRA P.R. *Steepest descent method with a generalized Armijo search for quasiconvex functions on riemannian manifolds*. *Journal of Mathematical Analysis and Applications*, Vol 341, pp. 467-477, 2008.

- [19] PAPA QUIROZ E.A. *Algumas Aplicações da Geometria Riemanniana à Otimização*. (COPPE/UFRJ, D.Sc., Engenharia de Sistemas e Computação). Tese - Universidade Federal do Rio de Janeiro, COPPE, Rio de Janeiro, 2007.
- [20] QUISPE E.M. *Método de Máximo Descenso en Variedades Riemannianas*. Tesis de Licenciatura. Facultad de Ciencias Naturales y Matematica, Universidad Nacional del Callao, Callao-Perú, 2008.
- [21] SEYMOUR LIPSCHUTZ. *Theory and problems of general topology*. McGraw-Hill, Inc. 1965.
- [22] SMITH S.T. *Optimization techniques on Riemannian manifolds*. In Hamiltonian and gradient flows, algorithms and control, volume 3 of Fields Inst. Commun., pages 113-136. Amer. Math. Soc., Providence, RI, 1994.
- [23] UDRISTE C. *Convex functions and optimization methods on riemannian manifolds*, volume 297 of Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht, 1994.
- [24] YANG W.H. and ZHANG L.H. *Optimality conditions of the nonlinear programming on Riemannian manifolds*. School of Mathematical Sciences, Fudan University, Shanghai, 200433, P. R. China, 2011.
- [25] YANG Y. *Globally convergent optimization algorithms on Riemannian manifolds: Uniform framework for unconstrained and constrained optimization*. J. Optim. Theory Appl., 132(2):245-265, 2007.

ANEXOS

■ Matriz de consistencia

Problema	Objetivos	Hipótesis	Metodología	Población
<p>Determinación del problema</p> <p>El modelo general que representa a un problema de optimización es:</p> $\begin{cases} \min f(x) \\ s.a. \\ x \in X \subseteq \mathbb{R}^n \end{cases} \quad (1)$ <p>El modelo (1) se puede generalizar para otros espacios más generales, por ejemplo si generalizamos para variedades obtendremos el modelo:</p> $\begin{cases} \min f(x) \\ s.a \\ x \in \mathcal{M} \end{cases} \quad (2)$ <p>donde, f es una función definida en una variedad \mathcal{M}.</p> <p>Para resolver el modelo (1) existen muchos métodos uno de ellos es el método de máximo descenso el cual genera la sucesión de puntos x_k dado por:</p> $x_{k+1} = x_k + \lambda_k d_k,$ <p>Si el método de máximo descenso se generaliza para resolver el problema (2), se denominará método de máximo descenso en variedades, y generará la siguiente sucesión de puntos x_k dado por:</p> $x_{k+1} = R_{x_k}(t_k \eta_k),$ <p>donde, $\eta_k \in T_{x_k} \mathcal{M}$, t_k es un escalar y R es una aplicación llamada retracción.</p> <p>Formulación del problema</p> <p>Se pretende resolver las siguientes interrogantes:</p> <ol style="list-style-type: none"> 1. ¿Será posible aplicar el método de máximo descenso, para resolver el problema (2)? 2. ¿Será posible implementar el método estudiado en MATLAB? 	<p>Objetivo General</p> <p>Estudiar las propiedades de convergencia del método de máximo descenso usando retracciones en variedades riemannianas, plantear su algoritmo e implementarlo en MATLAB.</p> <p>Objetivos Específicos</p> <ol style="list-style-type: none"> 1) Estudiar la aplicación retracción y determinarlos para algunas variedades particulares. 2) Estudiar la extensión del método de máximo descenso usando retracciones en variedades riemannianas. 3) Minimizar el cociente de Rayleigh sobre la esfera unitaria. 4) Minimizar la función de Brockett sobre la variedad de Stiefel. 	<p>Hipótesis General</p> <p>Conociendo los métodos clásicos de búsqueda lineal generalizado a variedades riemannianas, el algoritmo de dicho método se puede implementar en el software MATLAB de manera que se obtenga resultados muy cercanos al valor óptimo.</p> <p>Hipótesis específica</p> <p>Para implementar el algoritmo de nuestro método en estudio es necesario saber determinar la retracción de la variedad sobre la cual trabajamos, por tanto nosotros nos enfocamos en variedades matriciales de las cuales hallamos su retracción.</p>	<p>Tipo de investigación</p> <p>La investigación es de tipo teórico-aplicativa.</p> <p>Método</p> <p>la metodología usada es de tipo inductivo-deductivo.</p> <p>Diseño</p> <p>En la primera parte presentaremos la teoría básica de variedades diferenciables, de las cuales se estudia su topología, su diferenciabilidad entre otras propiedades y luego se da una introducción a las variedades riemannianas.</p> <p>En un segundo paso, definiremos la aplicación retracción, el cual es un paso clave para estudiar el método de máximo descenso, también se estudió la convergencia y velocidad de convergencia del método.</p> <p>Por último, aplicamos el método de máximo descenso para minimizar ejemplos específicos de los cuales implementamos su algoritmo y finalmente se presentan los resultados numéricos obtenidos en MATLAB.</p>	<p>Por ser nuestro trabajo netamente abstracto, no existe población que estudiar, sin embargo, nuestro estudio se encuentra inmerso dentro de los métodos de optimización en variedades riemannianas.</p>

■ Mapa conceptual del trabajo

