

UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”

DECANATO DE CIENCIAS Y TECNOLOGÍA

MAESTRÍA EN CIENCIAS - MENCIÓN OPTIMIZACIÓN

# Sobre Métodos de Optimización en Variedades.

AUTOR: LCDO. Luis Freytez

TUTOR: DR. Rómulo Castillo

Trabajo de grado presentado para optar al título de Magister Scientiae -

Mención Optimización

Barquisimeto, Septiembre 2013

# Índice general

---

<b>Introducción</b>	<b>1</b>
<b>1. Preliminares</b>	<b>4</b>
1.1. Superficies . . . . .	4
1.2. Variedades . . . . .	6
1.3. Aplicaciones diferenciables entre variedades . . . . .	7
1.4. Espacio Tangente . . . . .	8
1.5. Métricas Riemannianas . . . . .	10
1.5.1. Construcción de métricas riemannianas invariantes a través de H-traslación . . . . .	12
1.6. Campos de vectores, conexiones afines y derivada covariante . . . .	16
1.7. Curvatura de una variedad Riemanniana . . . . .	23
1.8. Preliminares de optimización . . . . .	29
1.9. Búsqueda lineal no monótona . . . . .	36
<b>2. Optimización en Variedades Riemannianas</b>	<b>38</b>

<i>ÍNDICE GENERAL</i>	II
2.1. Método de Gradiente Proyectado a lo Largo de Geodésicas. . . . .	38
2.2. Minimizando una Función Diferenciable Sobre una Variedad Dife- renciable. . . . .	41
2.3. Algoritmos Geodésicos en Geometria Riemanniana. . . . .	44
2.4. Una Regularización proximal del Método de descenso en Variedades Riemannianas. . . . .	46
2.5. Método de Descenso Para Minimización Cuasiconvexa en Varieda- des Riemannianas. . . . .	48
2.6. Un algoritmo de descenso para optimización cuasiconvexa en va- riedades Riemannias . . . . .	54
<b>3. Características de las métricas diagonales</b>	<b>57</b>
<b>4. Experimentos Numéricos</b>	<b>72</b>
4.1. Búsqueda de Armijo Generalizada . . . . .	73
4.2. Búsqueda no monótona . . . . .	75
<b>Bibliografía</b>	<b>89</b>

## Introducción

---

Una de las ramas de la matemática aplicada es la optimización, la cual consiste en maximizar o minimizar una función sujeta a algunas restricciones sobre su dominio. Sin perder generalidad, un problema de Optimización puede ser expresado, como:

$$\min_{s.a \ x \in \Omega} f(x), \quad (1)$$

en particular tenemos el problema siguiente:

$$\min_{s.a \ h(x)=0} f(x) \quad (2)$$

donde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  es una función de valores reales y  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $m < n$  es un función de valores en  $\mathbb{R}^m$ .

Existe una gran gama de métodos para resolver este tipo de problemas, entre los cuales, están los basados en proyección, entre ellos tenemos, el método del gradiente proyectado, el método de Newton proyectado, entre otros. Una desventaja del uso de estos métodos, es el alto costo computacional que se genera al proyectar los iterados sobre el conjunto factible; otra desventaja es la necesidad de la hipótesis de convexidad de la función  $f$ .

Considerando herramientas de la geometría riemanniana podemos suponer que 0 es valor regular del mapeo  $h$ , así se obtiene que  $M = \{x \in \mathbb{R}^n : h(x) = 0\}$  es una variedad riemanniana y el problema (2) se puede transformar en el siguiente problema irrestricto:

$$\underset{s.a. x \in M}{\text{mín}} f(x) \tag{3}$$

donde  $M$  es una variedad Riemanniana y  $f : M \rightarrow \mathbb{R}$  es una función de valores reales. Trabajando con este nuevo problema irrestricto ya no es necesario hacer proyecciones de los iterados en cada iteración. En adición, si introducimos una métrica riemanniana adecuada sobre la variedad  $M$  es posible convertir funciones no convexas a convexas.

El propósito de este trabajo es, en primer lugar, hacer una revisión de los principales métodos desarrollados hasta el presente, que usan la dirección de máximo descenso y variantes en la búsqueda de Armijo para resolver el problema de optimización considerado, usando elementos de geometría riemanniana a fin de obtener una buena comprensión de las ideas allí plasmadas. En segundo lugar, habiendo detallado el caso particular de las métricas diagonales, el proponer el

uso de búsquedas no monótonas en lugar de búsquedas de Armijo para lo cual se presentan ejemplos en donde su uso mejora el comportamiento numérico del algoritmo.

En el capítulo 1 de este trabajo, se mencionan algunos preliminares necesarios para entender conceptos básicos de la geometría riemanniana, en capítulo 2, hacemos un resumen detallado de algunos de los métodos desarrollados hasta ahora, para resolver problemas de optimización en variedades riemannianas, en capítulo 3, se estudian las características de las métricas riemannianas diagonales definidas en ortante positivo  $\mathbb{R}_{++}^n$  y el hipercubo abierto  $(0, 1)^n$ , se dan expresiones simples de los símbolos de Christoffel, ecuaciones geodésicas, gradiente y hessiano, herramientas que utilizamos en capítulo 4 para realizar experimentos numéricos de algunos de los métodos estudiados y de un método propuesto de optimización riemanniana, basado en la búsqueda no monótona de  $\mathbb{R}^n$ .

---

# Capítulo 1

## Preliminares

---

En este capítulo presentamos algunos preliminares de geometría riemanniana, extraídos en su gran mayoría de los libros de Do Carmo [7] y [8].

### 1.1. Superficies

**Definición 1.1.1** Sea  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  una función diferenciable definida en un abierto  $U$ . Definimos la **Diferencial** de  $F$  en el punto  $p \in U$  como una aplicación

$$dF_p : \mathbb{R}^n \rightarrow \mathbb{R}^m,$$

definida de la siguiente manera,  $dF_p(v) = \beta'(0)$ , con  $v \in \mathbb{R}^n$ , donde  $\beta = F \circ \alpha$  para alguna curva diferenciable  $\alpha : (-\varepsilon, \varepsilon) \rightarrow U$  tal que  $\alpha(0) = p$ ,  $\alpha'(0) = v$ . Así:

$$dF_p(v) = \beta'(0) = \left. \frac{d}{dt}(F \circ \alpha)(t) \right|_{t=0}.$$

**Definición 1.1.2** Sea  $F : \mathbb{R}^n \rightarrow \mathbb{R}^m$  una función diferenciable definida en un abierto  $U$ . Diremos que  $p \in U$  es punto crítico, si la diferencial de  $F$  en el punto

$p$ ,  $dF_p : \mathbb{R}^n \rightarrow \mathbb{R}^m$  no es sobreyectiva. La imagen  $F(p)$ , donde  $p$  es punto crítico es llamado valor crítico. Un punto de  $\mathbb{R}^m$  que no es valor crítico se llama valor regular de  $F$ , esto es,  $a \in F(U)$  es valor regular si  $dF_x$  es sobreyectiva para todo  $x \in F^{-1}(a)$ .

**Definición 1.1.3** (*Superficie Regular*). Un subconjunto  $S \subset \mathbb{R}^n$ , es una superficie regular de  $\mathbb{R}^n$  de dimensión  $k \leq n$  si para cada  $p$  en  $S$  existe una vecindad  $V$  de  $p$  en  $\mathbb{R}^n$ , un subconjunto abierto  $U$  de  $\mathbb{R}^k$  y una aplicación  $\mathcal{X} : U \rightarrow V \cap S$  que cumple:

(i)  $\mathcal{X}$  es diferenciable en  $U$ .

(ii)  $\mathcal{X}$  es un homeomorfismo.

(iii) Para cada  $p$  en  $U$  la diferencial  $d\mathcal{X}_p : \mathbb{R}^k \rightarrow \mathbb{R}^n$  es Inyectiva.

A la aplicación  $\mathcal{X} : U \rightarrow V \cap S$  se le llama una parametrización, carta ó sistema de coordenadas locales en una vecindad de  $p$ .  $S \cap U$  es llamada vecindad coordenada de  $p$ .

**Proposición 1.1.1** Sea  $U$  un subconjunto abierto de  $\mathbb{R}^n$  y  $F : U \subset \rightarrow \mathbb{R}^m$  una función diferenciable en  $U$  con valor regular  $a \in \mathbb{R}^m$ , entonces  $F^{-1}(a)$  es superficie regular de dimensión  $n - m$ .

**Proposición 1.1.2** Sea  $U$  un subconjunto abierto de  $\mathbb{R}^n$  y  $f : U \subset \rightarrow \mathbb{R}$  una



función diferenciable tal que  $\nabla f(x) \neq 0$  para todo  $x \in f^{-1}(a)$ , entonces  $S = f^{-1}(a)$  es superficie regular.

**Proposición 1.1.3** (Cambio de coordenadas). Sean  $\mathcal{X} \rightarrow S$  y  $\mathcal{Y} \rightarrow S$  dos parametrizaciones de  $S$  tal que  $W = \mathcal{X}(U) \cap \mathcal{Y}(V) \neq \emptyset$ , entonces la aplicación  $h = \mathcal{Y}^{-1} \circ \mathcal{X} : \mathcal{X}^{-1}(W) \rightarrow \mathcal{Y}^{-1}(W)$  es llamada cambio de coordenadas.

**Proposición 1.1.4** Sea  $S$  una superficie regular de  $\mathbb{R}^n$  de dimensión  $k$ . El cambio de coordenadas  $\mathcal{Y}^{-1} \circ \mathcal{X} : \mathcal{X}^{-1}(W) \rightarrow \mathcal{Y}^{-1}(W)$  es un difeomorfismo.

## 1.2. Variedades

**Definición 1.2.1** Una variedad diferenciable de dimensión  $n$ , es un conjunto  $M$  y una familia de aplicaciones inyectivas  $\mathcal{X}_\alpha : U_\alpha \rightarrow M$ ,  $\alpha \in I$ , definidas en abiertos  $U_\alpha$  de  $\mathbb{R}^n$  en  $M$  tales que cumplen las siguientes condiciones:

- $M = \bigcup_{\alpha \in I} \mathcal{X}_\alpha(U_\alpha)$
- Para todo par  $\mathcal{X}_\alpha$  y  $\mathcal{X}_\beta$  con  $\mathcal{X}_\alpha(U_\alpha) \cap \mathcal{X}_\beta(U_\beta) = W \neq \emptyset$ , los conjuntos  $\mathcal{X}_\alpha^{-1}(W)$  y  $\mathcal{X}_\beta^{-1}(W)$  son abiertos en  $\mathbb{R}^n$  y las aplicaciones  $\mathcal{X}_\beta^{-1} \circ \mathcal{X}_\alpha : \mathcal{X}_\alpha^{-1}(W) \rightarrow \mathcal{X}_\beta^{-1}(W)$  son diferenciables.

El par  $(U_\alpha, \mathcal{X}_\alpha)$  con  $p \in \mathcal{X}_\alpha(U_\alpha)$  es llamado una parametrización de  $M$  de  $p$ ,  $\mathcal{X}_\alpha(U_\alpha)$  es llamada vecindad coordenada de  $p$ . Una familia  $\{(U_\alpha, \mathcal{X}_\alpha)\}$  satisfaciendo los items 1 y 2 es llamada estructura diferenciable de  $M$ .

**Proposición 1.2.1** *Toda superficie regular de  $\mathbb{R}^n$  de dimensión  $k$  es una variedad diferenciable de la misma dimensión.*

### 1.3. Aplicaciones diferenciables entre variedades

**Definición 1.3.1** *Sea  $f : U \subset M \rightarrow \mathbb{R}$  una función definida en un subconjunto abierto  $U$  de una variedad diferenciable  $M$ . Diremos que  $f$  es diferenciable en  $p \in U$ , si para alguna parametrización  $\mathcal{X}_\alpha : U_\alpha \subset \mathbb{R}^n \rightarrow M$ , con  $p \in \mathcal{X}_\alpha(U_\alpha) \subset U$ , la composición  $f \circ \mathcal{X}_\alpha : U_\alpha \subset \mathbb{R}^n \rightarrow \mathbb{R}$  es diferenciable en  $\mathcal{X}_\alpha^{-1}(p)$ . Se dice que  $f$  es diferenciable en  $U$  si es diferenciable en todo punto de  $U$ .*

**Definición 1.3.2** *Una curva sobre una variedad diferenciable  $M$  es una función  $\gamma : I \rightarrow M$  donde  $I = (-\varepsilon, \varepsilon)$ . Diremos que  $\gamma$  es diferenciable en  $t_0 \in I$  si para alguna parametrización  $\mathcal{X}_\alpha : U_\alpha \subset \mathbb{R}^n \rightarrow M$  con  $\gamma(t_0) \in \mathcal{X}_\alpha(U_\alpha)$ , la compuesta  $\mathcal{X}_\alpha^{-1} \circ \gamma : I \rightarrow U_\alpha$  es diferenciable en  $t_0$ , donde  $\gamma(I) \subset \mathcal{X}_\alpha(U_\alpha)$ . Se dice que  $\gamma$  es diferenciable en  $I$  si es diferenciable en todo  $t \in I$ .*

**Definición 1.3.3** Sean  $M_1$  y  $M_2$  variedades diferenciables de dimensión  $m$  y  $n$  respectivamente. Una aplicación  $\varphi : M_1 \rightarrow M_2$  es diferenciable en  $p \in V$ , si dadas  $\mathcal{X}_1 : U_1 \subset \mathbb{R}^n \rightarrow M_1$  parametrización de  $M_1$  en  $p$  y  $\mathcal{X}_2 : U_2 \subset \mathbb{R}^m \rightarrow M_2$  parametrización de  $M_2$  en  $\varphi(p)$ , con  $\varphi(\mathcal{X}_1(U_1)) \subset \mathcal{X}_2(U_2)$ , la aplicación  $\mathcal{X}_2^{-1} \circ \varphi \circ \mathcal{X}_1 : U_1 \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$  es diferenciable en  $\mathcal{X}_1^{-1}(p)$

**Definición 1.3.4** Sea Una aplicación  $\varphi : M_1 \rightarrow M_2$  es diferenciable entre dos variedades diferenciables. Decimos que  $\varphi$  es difeomorfismo si  $\varphi$  es biyectiva y  $\varphi^{-1}$  es diferenciable.

## 1.4. Espacio Tangente

**Definición 1.4.1** Sea  $M$  una variedad diferenciable. Consideremos una curva diferenciable  $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ , donde  $\gamma(0) = p$  y sea  $D_p$  el conjunto de las funciones de  $M$  diferenciables en  $p$ ,  $D_p = \{f : M \rightarrow \mathbb{R} / f \text{ es diferenciable en } p\}$ ; se define el **Vector Tangente** a la curva  $\gamma$  en  $t = 0$  como la función  $\gamma'(0) : D_p \rightarrow \mathbb{R}$  dada por :

$$\gamma'(0)f = \left. \frac{d(f \circ \gamma)}{dt} \right|_{t=0}, \quad f \in D_p. \quad (1.1)$$

**Definición 1.4.2** (Espacio tangente). El espacio tangente a una variedad  $M$  en un punto  $p$  representado por  $T_p M$ , es el conjunto de todos los vectores tangentes a  $M$  en  $p$ . Así,  $T_p M = \{v \in \mathbb{R}^m : v \text{ es un vector tangente en } p\}$ .

**Observación 1.4.1** Si escogemos una parametrización  $\mathcal{X}_\alpha : U \subset \mathbb{R}^n \rightarrow M$  con  $p = \mathcal{X}(0)$  y  $q \in U$  podemos restringir la función  $f$  y la curva  $\gamma$  en esta parametrización por :

$$f \circ \mathcal{X}(q) = f(q) = f(q_1, \dots, q_n) \text{ y}$$

$$\mathcal{X}^{-1} \circ \gamma(t) = (q_1(t), \dots, q_n(t)),$$

restringiendo  $f$  a  $\gamma$ , tenemos

$$\begin{aligned} \gamma'(0)f &= \left. \frac{d}{dt}(f \circ \gamma) \right|_{t=0} = \left. \frac{d}{dt}f(q_1(t), \dots, q_n(t)) \right|_{t=0} \\ &= \sum_{i=1}^n q'_i(0) \left( \frac{\partial f}{\partial q_i}(p) \right) = \left( \sum_{i=1}^n q'_i(0) \left( \frac{\partial}{\partial q_i} \right)_0 \right) f. \end{aligned}$$

Así,

$$\gamma'(0) = \sum_{i=1}^n q'_i(0) \left( \frac{\partial}{\partial q_i} \right)_0 \quad (1.2)$$

es la expresión del vector tangente a  $f$  en  $p$  con relación a la parametrización  $\mathcal{X}$ .

**Observación 1.4.2** La expresión (1.2) muestra que el vector tangente a una curva  $\gamma$  en  $p$  depende de las derivadas de  $\gamma$  en un sistema de coordenadas. De la elección de una parametrización obtenemos  $n$  vectores  $\left\{ \left( \frac{\partial}{\partial x_1} \right), \dots, \left( \frac{\partial}{\partial x_n} \right) \right\}$  en  $T_p M$  que generan, por (1.2), los vectores en  $T_p M$ .

**Proposición 1.4.1** El espacio tangente de una variedad diferenciable que es un subconjunto abierto de  $\mathbb{R}^n$  es el propio  $\mathbb{R}^n$ .

**Definición 1.4.3** Sean  $M_1$  y  $M_2$  dos variedades diferenciables de dimensión  $n$  y  $m$  respectivamente y sea  $\varphi : M_1 \rightarrow M_2$  una aplicación diferenciable. Para cada  $p \in M_1$  y cada  $v \in T_p M_1$ , escojamos una curva diferenciable  $\alpha : (-\varepsilon, \varepsilon) \rightarrow M_1$ , con  $\alpha(0) = p$  y  $\alpha'(0) = v$ . Definiendo  $\beta = \varphi \circ \alpha$  la aplicación:

$$d\varphi_p : T_p M_1 \rightarrow T_{\varphi(p)} M_2,$$

dada por  $d\varphi_p(v) = \beta'(0)$  es una aplicación lineal que no depende de la elección  $\alpha$ . Esta aplicación es llamada la diferencial de  $\varphi$  en  $p$ .

## 1.5. Métricas Riemannianas

**Definición 1.5.1** Sea  $M$  una variedad diferenciable. Una métrica riemanniana es una aplicación que asocia a cada  $p \in M$  un producto interno (bilineal, simétrico y definido positivo)  $\langle \cdot, \cdot \rangle_p$ , dado por :

$$\langle \cdot, \cdot \rangle_p = T_p M \times T_p M \rightarrow \mathbb{R},$$

que varía diferenciablemente en el siguiente sentido: si  $\mathcal{X} : U \subset \mathbb{R}^n \rightarrow M$  es un sistema de coordenadas locales en torno de  $p$ , con  $\mathcal{X}(x_1, x_2, x_3, \dots, x_n) = q \in \mathcal{X}(U)$  y  $\frac{\partial}{\partial x_i}(q) = d\mathcal{X}_p(0, 0, \dots, 0, 1, 0, \dots, 0, 0)$ , entonces la función:  $g_{ij} : U \rightarrow \mathbb{R}$  definida por:

$$g_{ij}(x_1, x_2, \dots, x_n) = \left\langle \frac{\partial}{\partial x_i}(q), \frac{\partial}{\partial x_j}(q) \right\rangle_q,$$

es una función diferenciable en  $U$ .

Las funciones  $g_{ij}$  son llamadas expresiones de la métrica riemanniana en el sistema coordenado  $\mathcal{X}$  y la matriz  $G = (g_{ij})$  es la representación de la métrica riemanniana.

**Definición 1.5.2** (*Variedad riemanniana*). Una variedad diferenciable para la cual se define una métrica riemanniana se denomina una variedad riemanniana.

**Definición 1.5.3** El Producto interno de dos vectores  $u, v \in T_p M$  es definido por  $\langle u, v \rangle_p = g_x(u, v)$ , donde  $g_x$  es la métrica riemanniana evaluada en el punto  $x$ . La norma de un vector  $v \in T_p M$  es  $\|v\|_p = \sqrt{\langle v, v \rangle_p}$

**Definición 1.5.4** Un grupo de Lie es una variedad diferenciable  $M$  con una estructura de grupo  $\bullet$ , tal que la aplicación  $M \times M \rightarrow M$  dada por  $(x, y) \rightarrow x \bullet y^{-1}$  es diferenciable. Para un elemento  $y \in M$  la traslación a izquierda por  $x$  es el mapeo  $L_x : M \rightarrow M$  definido por  $L_x(y) = x \bullet y$ , el cual es un difeomorfismo.

**Definición 1.5.5** Sea  $M$  un grupo de Lie. Una métrica riemanniana  $G$  es en  $M$  es invariante por la izquierda si

$$\langle u, v \rangle_y = \langle d(L_x)u, d(L_x)v \rangle, \text{ para todo } x, y \in M \text{ y } u, v \in T_y M$$

### 1.5.1 Construcción de métricas riemannianas invariantes a través de H-traslación

Sea  $M$  una variedad diferenciable,  $x \in M$  y  $H : M \rightarrow M$  un mapeo diferenciable. Para construir una métrica riemanniana invariante a través de H-traslación sigamos los siguientes pasos (Do Carmo 36):

1. Definamos en  $M$  una estructura de grupo de Lie.
2. Dado un  $x \in M$ , consideremos el elemento

$$\langle, \rangle_{L_{H(x)^{-1}x}} : T_{L_{H(x)^{-1}x}}M \times T_{L_{H(x)^{-1}x}}M \rightarrow \mathbb{R}$$

3.  $\forall u, v \in T_x M$  defina la métrica

$$\langle u, v \rangle_x = \langle d(L_{H(x)^{-1}})_x u, d(L_{H(x)^{-1}})_x v \rangle_{L_{H(x)^{-1}x}} \quad (1.3)$$

El algoritmo anterior se aplica para generar métricas riemannianas diagonales asociadas para algunos conjuntos naturales viables para los problemas de optimización lineal.

#### **Ejemplo 1.5.1 Métricas diagonales en $\mathbb{R}_{++}^n$ .**

Consideremos la variedad diferenciable  $\mathbb{R}_{++}^n$  y sea  $H: \mathbb{R}_{++}^n \rightarrow \mathbb{R}_{++}^n$  una función tal que  $H(x) = (h_1(x_1), h_2(x_2), \dots, h_n(x_n))$ , donde  $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$  son funciones diferenciables.

1. Definamos en  $M$  una estructura de grupo de Lie.

Sea  $x = (x_1, x_2, \dots, x_n)$  y  $y = (y_1, y_2, \dots, y_n) \in \mathbb{R}_{++}^n$ , definamos

$$x \bullet y = (x_1 \cdot y_1, x_2 \cdot y_2, \dots, x_n \cdot y_n).$$

Esa operación define una estructura de grupo con elemento identidad  $e = (1, 1, \dots, 1)$  y el inverso de un elemento  $x = (x_1, x_2, \dots, x_n)$  dado por  $x^{-1} = (\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_n})$ .

2. Definamos el producto interno en  $L_{H(x)^{-1}x}$ .

Sea  $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}_{++}^n$ , entonces

$$L_{H(x)^{-1}x} = H(x)^{-1} \bullet x = \left( \frac{x_1}{h_1(x_1)}, \frac{x_2}{h_2(x_2)}, \dots, \frac{x_n}{h_n(x_n)} \right).$$

Como  $T_{H(x)^{-1} \bullet x} \mathbb{R}_{++}^n = \mathbb{R}^n$ , podemos definir un producto interno en este punto como un producto escalar euclidiano

$$\langle v, w \rangle_{L_{H(x)^{-1}x}} = (v, w) = v^T w.$$

3. Definamos la métrica para todo  $x \in \mathbb{R}_{++}^n$ .

En primer lugar, vamos a obtener  $d(L_{H(x)^{-1}x})$ . La aplicación  $L_{H(x)} : \mathbb{R}_{++}^n \rightarrow$

$\mathbb{R}_{++}^n$  es por definición  $L_{H(x)}y = H(x) \bullet y = (h_1(x_1) \cdot y_1, h_2(x_2) \cdot y_2, \dots, h_n(x_n) \cdot y_n)$ .

Entonces  $d(L_{H(x)})y : \mathbb{R}^n \rightarrow \mathbb{R}^n$  puede ser expresada por la matriz

$$d(L_{H(x)})y = \text{diag}(h_1(x_1), h_2(x_2), \dots, h_n(x_n))$$



y, en consecuencia

$$d(L_{H(x)^{-1}})y = \text{diag}\left(\frac{1}{h_1(x_1)}, \frac{1}{h_2(x_2)}, \dots, \frac{1}{h_n(x_n)}\right).$$

Ahora,  $\forall u, v \in T_x \mathbb{R}_{++}^n = \mathbb{R}^n$  definimos

$$\langle u, v \rangle_x = \langle d(L_{H(x)^{-1}})_x u, d(L_{H(x)^{-1}})_x v \rangle_{L_{H(x)^{-1}}x} = \sum_{i=1}^n \frac{u_i v_i}{h_i(x_i)^2} = u^T G(x) v, \quad (1.4)$$

donde,

$$G(x) = \text{diag}\left(\frac{1}{h_1^2(x_1)}, \frac{1}{h_2^2(x_2)}, \dots, \frac{1}{h_n^2(x_n)}\right).$$

Así se obtiene que,

$$g_{ij}(x) = \langle e_i, e_j \rangle_x = \frac{\delta_{ij}}{h_i(x_i)h_j(x_j)}.$$

En particular:

- Si  $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ ,  $h_i(x_i) = x_i^{\frac{r}{2}}$ , donde  $r$  es un escalar distinto de cero,

entonces:

$$G(x) = \text{diag}\left(\frac{1}{x_1^r}, \frac{1}{x_2^r}, \dots, \frac{1}{x_n^r}\right) = X^{-r},$$

donde  $X = \text{diag}(x_1, x_2, \dots, x_n)$ .

- Si  $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ ,  $h_i(x_i) = s_i^{-\frac{r}{2}} x_i^{\frac{r}{2}}$ , donde  $s_i \in \mathbb{R}_{++}$  es fijo y  $r$  es un

escalar distinto de cero, entonces:

$$G(x) = \text{diag}\left(\frac{s_1^r}{x_1^r}, \frac{s_2^r}{x_2^r}, \dots, \frac{s_n^r}{x_n^r}\right) = S^r X^{-r},$$

donde  $X = \text{diag}(x_1, x_2, \dots, x_n)$ .

**Ejemplo 1.5.2 Métrica riemanniana diagonal en el hipercubo.**

Consideremos la subvariedad diferencial  $\mathcal{C}_0^n = (0, 1)^n$  de  $\mathbb{R}_{++}^n$  y sea  $H : \mathcal{C}_0^n \rightarrow \mathbb{R}_{++}^n$ , tal que  $H(x) = (h_1(x_1), h_2(x_2), \dots, h_n(x_n))$ , donde  $h_i : \mathcal{C}_0^n \rightarrow \mathbb{R}_{++}$  son funciones diferenciables. Ahora, podemos definir la métrica en  $\mathcal{C}_0^n$  como la métrica inducida de  $\mathbb{R}_{++}^n$ , introducida en el ejemplo anterior. Así,  $\forall u, v \in T_x \mathcal{C}_0^n = \mathbb{R}^n$

$$\langle u, v \rangle_x = \sum_{i=1}^n \frac{u_i v_i}{h_i(x_i)^2} = u^T G(x) v, \quad (1.5)$$

donde,

$$G(x) = \text{diag} \left( \frac{1}{h_1^2(x_1)}, \frac{1}{h_2^2(x_2)}, \dots, \frac{1}{h_n^2(x_n)} \right).$$

En particular:

- Si  $h_i(x_i) = (x_i^r(1-x_i)^r / (x_i^r + (1-x_i)^r))^{1/2}$ , para  $x_i \in (0, 1)$ ,  $i = 1, 2, \dots, n$  y  $r$  es un escalar distinto de cero, entonces:

$$G(x) = X^{-r} + (I - X)^{-r}.$$

- Si  $h_i(x_i) = \sin^2(\pi x_i) / \pi$ , para  $x_i \in (0, 1)$ ,  $i = 1, 2, \dots, n$ , tenemos la métrica trigonométrica:

$$\pi^2 \csc^4(\pi x) = \text{diag}(\pi^2 \csc^4(\pi x_1), \pi^2 \csc^4(\pi x_2), \dots, \pi^2 \csc^4(\pi x_n))$$

- Si  $h_i(x_i) = x_i^{r/2}(1 - x_i)^{r/2}$ , para  $x_i \in (0, 1)$ ,  $i = 1, 2, \dots, n$  y  $r$  es un escalar distinto de cero, entonces:

$$G(x) = X^{-r}(I - X)^{-r}.$$

## 1.6. Campos de vectores, conexiones afines y derivada covariante

**Definición 1.6.1** (*Campo de vectores en una variedad diferenciable*). Un campo de vectores  $X$  en una variedad diferenciable  $M$  es una correspondencia que a cada punto  $p \in M$  asocia un vector  $X(p) \in T_p M$ .

Considerando una parametrización  $\mathcal{X} : U \subset \mathbb{R}^n \rightarrow M$ , es posible escribir:

$$X(p) = \sum_{i=1}^n a_i(p) \left( \frac{\partial}{\partial x_i} \right)_p,$$

donde cada  $a_i : M \rightarrow \mathbb{R}$  es una función en  $M$  y  $\left\{ \left( \frac{\partial}{\partial x_i} \right)_p \right\}$  es una base asociada a  $\mathcal{X}$ ,  $1 \leq i \leq n$ . Diremos que  $X$  es diferenciable si y solo si las funciones  $a_i$  son diferenciables.

**Definición 1.6.2** (*Campo de vectores a lo largo de curvas*). Un campo vectorial  $V$  a lo largo de una curva  $\alpha : I \rightarrow M$  es una aplicación que a cada  $\alpha(t) \in M$  asocia un vector tangente  $V(t) \in T_{\alpha(t)} M$ .

**Conexiones Afines.**

Denotemos  $TM$  como el conjunto de espacios tangentes definidos en  $M$ . Sea  $\mathcal{H} = \mathcal{H}(M) = \{X : M \rightarrow TM : \text{para cada } p \in M, X(p) \in T_pM, \text{ y } X \in C^\infty\}$  el conjunto de campo de vectores de clase  $C^\infty$  y  $D = D(M) = \{f : M \rightarrow \mathbb{R} : f \in C^\infty\}$  el conjunto de funciones reales de clase  $C^\infty$ .

**Definición 1.6.3** Una conexión afín es una aplicación  $\nabla : \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{H}$  donde a cada par de campos  $(X, Y)$  se asocia otro campo  $\nabla_X Y$  tal que para todo  $X, Y, Z \in \mathcal{H}$ , y  $f, g \in D$  se cumple:

$$\text{i) } \nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z;$$

$$\text{ii) } \nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z;$$

$$\text{iii) } \nabla_X fY = f\nabla_X Y + X(f)Y, \text{ donde } X(f) = \sum_{i=1}^n a_i(\cdot) \frac{\partial f(\cdot)}{\partial x_i}$$

**Proposición 1.6.1** (Derivada Covariante) Sea  $M$  una variedad diferenciable con una conexión afín  $\nabla$ . Entonces existe una única aplicación que asocia a un campo vectorial  $V$  a lo largo de una curva diferenciable  $\alpha : I \rightarrow M$  otro campo vectorial  $\frac{DV}{dt}$  a lo largo de  $\alpha$ , denominado derivada covariante de  $V$  a lo largo de  $\alpha$ , tal que para todo  $V, W$  campo de vectores a lo largo de  $\alpha$  y  $f : I \rightarrow \mathbb{R}$  una función diferenciable en  $I$  se cumple:

$$\text{i) } \frac{D}{dt}(V + W) = \frac{DV}{dt} + \frac{DW}{dt}.$$

$$\text{ii) } \frac{D}{dt}(fV) = \frac{df}{dt}V + f\frac{DV}{dt}.$$

$$\text{iii) } \text{si } V(t) = Y(\alpha(t)), \text{ donde } Y \in \mathcal{H}, \text{ entonces } \frac{DV}{dt} = \nabla_{\frac{d\alpha}{dt}}Y.$$

**Expresión de la conexión afín en términos de las coordenadas locales.**

Suponga que los campos de vectores  $X, Y \in \mathcal{H}$  son representados en una cierta vecindad local  $\mathcal{X} : U \subset \mathbb{R}^n \rightarrow M$  de un punto  $p$ , por:

$$X = \sum_{i=1}^m x_i X_i, Y = \sum_{i=1}^m y_i X_i,$$

donde  $X_i = \frac{\partial}{\partial x_i}$  representan los vectores de la base del sistema de coordenadas locales.

Usando las propiedades de la definición de la conexión afín:

$$\begin{aligned} \nabla_X Y &= \nabla_{\sum x_i X_i} \left[ \sum_j y_j X_j \right] = \sum_i x_i \left[ \nabla_{X_i} \left( \sum_j y_j X_j \right) \right] \\ &= \sum_i x_i \left[ \sum_j (y_j \nabla_{X_i} X_j) \right] + \sum_i x_i \left[ \sum_j \left( \frac{\partial y_j}{\partial x_i} X_j \right) \right], \end{aligned}$$

escribiendo  $\nabla_{X_i} X_j$  en función de la base local:

$$\nabla_{X_i} X_j = \sum_{k=1}^n \Gamma_{ij}^k X_k, \tag{1.6}$$

sustituyendo en la ecuación anterior obtenemos:

$$\nabla_X Y = \sum_{k=1}^n \left( \sum_{i,j=1}^n x_i y_j \Gamma_{ij}^k + \sum_i x_i \frac{\partial y_k}{\partial x_i} \right) X_k.$$

**Definición 1.6.4** *Los símbolos de Christoffel, o coeficientes de la conexión afín  $\nabla$  en  $U$ , son las funciones diferenciables  $\Gamma_{ij}^k : U \subset M \rightarrow R$  definidas por (1.6)*

**Expresión de la derivada covariante en términos de las coordenadas locales y de los símbolos de Christoffel.**

Sea  $\mathcal{X} : U \rightarrow M$  un sistema de coordenadas locales en torno de  $p \in M$ . Un resultado obtenido al demostrar la Proposición (1.6.1) es:

$$\frac{DV}{dt} = \sum_{j=1}^n \frac{dv^j}{dt} X_j + \sum_{i,j=1}^n v^j \frac{dx_i}{dt} \nabla_{X_i} X_j,$$

sustituyendo  $\nabla_{X_i} X_j = \sum_{k=1}^n \Gamma_{ij}^k X_k$  en la ecuación anterior tenemos

$$\begin{aligned} \frac{DV}{dt} &= \sum_{j=1}^n \frac{dv^j}{dt} X_j + \sum_{i,j=1}^n v^j \frac{dx_i}{dt} \left( \sum_{k=1}^n \Gamma_{ij}^k X_k \right), \\ &= \sum_{j=1}^n \frac{dv^j}{dt} X_j + \sum_{k=1}^n \sum_{i,j=1}^n v^j \frac{dx_i}{dt} \Gamma_{ij}^k X_k \end{aligned}$$

y así,

$$\frac{DV}{dt} = \sum_{k=1}^n \left( \frac{dv^k}{dt} + \sum_{i,j=1}^n v^j \frac{dx_i}{dt} \Gamma_{ij}^k \right) X_k, \quad (1.7)$$

es la expresión de la derivada covariante en términos de coordenadas locales y de los símbolos de Christoffel.

**Definición 1.6.5 (Geodésicas)** *Una curva parametrizada  $\alpha : I \rightarrow M$  es una geodésica si el campo tangente  $\frac{d\alpha}{dt}$  verifica:*

$$\frac{D}{dt} \left( \frac{d\alpha}{dt} \right) = 0 \quad (1.8)$$

**Definición 1.6.6** (*Campos paralelos*) Dado  $M$  una variedad diferenciable, una conexión afín  $\nabla$  y un campo  $V$  a lo largo de una curva diferenciable  $\alpha : I \rightarrow M$ ,  $V$  es denominado campo paralelo si

$$\frac{DV}{dt} = 0, \forall t \in I \quad (1.9)$$

Así, si  $\alpha$  es una geodésica, entonces  $\frac{d\alpha}{dt}$  es paralelo.

**Proposición 1.6.2** Sea  $M$  una variedad diferenciable con una conexión afín  $\nabla$ .

Sea  $\alpha : I \rightarrow M$  una curva diferenciable en  $M$  y  $V_0$  un vector tangente a  $M$  en  $\alpha(t_0)$ ,  $t_0 \in I$ . Entonces existe un único campo de vectores paralelo  $V$  a lo largo de  $\alpha$ , tal que  $V(t_0) = V_0$  ( $V(t)$  es llamado transporte paralelo de  $V(t_0)$  a lo largo de  $\alpha$ ).

### Ecuaciones Geodésicas

De la expresión (1.7), un campo paralelo  $V$  es determinado por las ecuaciones

$$\sum_{k=1}^n \left( \frac{dv^k}{dt} + \sum_{i,j=1}^n v^j \frac{dx_i}{dt} \Gamma_{ij}^k \right) X_k = 0,$$

de forma equivalente,

$$\frac{dv^k}{dt} + \sum_{i,j=1}^n v^j \frac{dx_i}{dt} \Gamma_{ij}^k = 0, \quad k = 1, 2, \dots, n.$$

Cuando se trata de una geodésica  $x(t) = (x_1(t), \dots, x_n(t))$ , se tiene  $v^j = \frac{dx_j}{dt}$ , entonces esta última ecuación se transforma en:

$$\frac{d^2 x_k}{dt^2} + \sum_{i,j=1}^n \Gamma_{ij}^k \frac{dx_j}{dt} \frac{dx_i}{dt} = 0, \quad k = 1, 2, \dots, n, \quad (1.10)$$

el cual es un sistema de  $n$  ecuaciones diferenciales de 2do. orden, que posee solución única en algún intervalo  $I = [a, b]$ , verificando  $x(0) = \alpha(0) = p$  y  $\frac{dx}{dt}(0) = \alpha'(0) = v$ .

**Proposición 1.6.3** *Dado un punto  $p \in M$  y un vector  $v \in T_p M$ ,  $v \neq 0$ , entonces existe un  $\varepsilon > 0$  y una única geodésica parametrizada  $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$  tal que  $\gamma(0) = p$ ,  $\gamma'(0) = v$ . Notación  $\gamma(t, v) = \gamma(t) = \gamma$ .*

**Definición 1.6.7** *Dado  $p \in M$  y  $v \in T_p M$ ,  $v \neq 0$ , sabemos que existe una única geodésica  $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$  tal que  $\gamma(0) = p$  y  $\gamma'(0) = v$ , entonces se define la aplicación exponencial,  $\exp_p : T_p M \rightarrow M$ , como:  $\exp_p(v) = \gamma(1, v) = \gamma(1)$  y  $\exp_p(0) = p$*

### Conexión afín en variedades riemannianas

**Definición 1.6.8** *Sea  $M$  una variedad diferenciable con una conexión afín  $\nabla$  y una métrica riemanniana  $\langle \cdot, \cdot \rangle$ . Se dice que  $\nabla$  es compatible con la métrica  $\langle \cdot, \cdot \rangle$ , si para todo par de campos de vectores  $V$  y  $W$  a lo largo de la curva diferenciable  $\alpha : I \rightarrow M$  se tiene:*

$$\frac{d}{dt} \langle V, W \rangle = \left\langle \frac{DV}{dt}, W \right\rangle + \left\langle V, \frac{DW}{dt} \right\rangle \quad (1.11)$$



**Proposición 1.6.1** *Si la conexión afín  $\nabla$  es compatible con  $\langle, \rangle$  y  $V, W$  son campos paralelos a lo largo de una curva diferenciable  $\alpha : I \rightarrow M$  entonces,  $\langle V, W \rangle$  es constante. En particular si  $\alpha(t) = (\alpha_1(t), \dots, \alpha_1(t))$  es una geodésica,  $\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \rangle$  es constante.*

**Definición 1.6.9** *Una conexión afín  $\nabla$  en una variedad riemanniana  $M$  es compatible con  $\langle, \rangle$  si y solamente si:*

$$X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle, \quad \forall X, Y, Z \in \mathcal{H}.$$

**Definición 1.6.10** *Una conexión afín  $\nabla$  en una variedad riemanniana  $M$  es llamada simétrica si:*

$$\nabla_X Y - \nabla_Y X = [X, Y] \quad \forall X, Y \in \mathcal{H}, \text{ donde } [X, Y] = XY - YX.$$

**Definición 1.6.11** *(Levi-Civita). Dada una variedad riemanniana  $M$ , existe una única conexión afín  $\nabla$  en  $M$  satisfaciendo las condiciones:*

- a)  $\nabla$  es simétrica.
- b)  $\nabla$  es compatible con la métrica riemanniana.

*(Esta conexión es denominada conexión riemanniana).*

**Definición 1.6.12** *Dado un sistema de coordenadas  $(U, \mathcal{X})$ , las funciones conocidas como símbolos de Christoffel  $\Gamma_{ij}^k : U \subset M \rightarrow \mathbb{R}$  que definen los coeficientes de*

conexión  $\nabla_{X_i} X_j = \sum_{k=1}^n \Gamma_{ij}^k X_k$ , están relacionados con la métrica de la siguiente manera :

$$\Gamma_{ij}^m = \frac{1}{2} \sum_k^n \left\{ \frac{\partial}{\partial x_i} g_{jk} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right\} g^{km}, \quad (1.12)$$

donde  $g_{ij} = \langle \frac{\partial}{\partial x_i}, \frac{\partial}{\partial x_j} \rangle$  son elementos de la matriz  $G(x)$  y  $g^{ij}$  los elementos de la inversa  $G(x)^{-1}$ .

## 1.7. Curvatura de una variedad Riemanniana

**Definición 1.7.1** Una curvatura  $K$  de una variedad Riemanniana  $M$  es una correspondencia que asocia a cada par  $X, Y \in \mathcal{H}$  una aplicación  $K(X, Y) : \mathcal{H} \rightarrow \mathcal{H}$  dada por :

$$k(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X, Y]} Z, \quad Z \in \mathcal{H},$$

donde  $\nabla$  es una conexión de Riemanniana de  $M$ .

**Observación 1.7.1** Si consideramos un sistema de coordenadas  $(U, \mathcal{X})$  en torno del punto  $p$  y  $\{X_i\}$ ,  $i = 1, 2, \dots, n$  es una base de  $T_p M$  obtenemos:

$$K(X_i, X_j)X_k = (\nabla_{X_j} \nabla_{X_i} - \nabla_{X_i} \nabla_{X_j})X_k$$

**Proposición 1.7.1** la curvatura  $K$  de una variedad Riemanniana cumple las siguientes propiedades:

i)  $K$  es bilineal en  $\mathcal{H} \times \mathcal{H}$ , esto es,

$$K(fX_1 + gX_2, Y_1) = fK(X_1, Y_1) + gK(X_2, Y_1),$$

$$K(X_1, fY_1 + gY_2) = fK(X_1, Y_1) + gK(X_1, Y_2),$$

donde  $f, g \in C^\infty(M)$ ,  $X_1, X_2, Y_1, Y_2 \in \mathcal{H}$ .

ii) Para todo  $X, Y \in \mathcal{H}$  el operador curvatura  $K(X, Y)$  es lineal, esto es,

$$K(X, Y)(Z + W) = K(X, Y)Z + K(X, Y)W,$$

$$K(X, Y)fZ = fK(X, Y)Z,$$

donde  $f \in C^\infty(M)$ ,  $Z, W \in \mathcal{H}$ .

iii) La curvatura es antisimétrica, esto es ,

$$K(X, Y) = -K(Y, X).$$

**Proposición 1.7.2** Sea  $(U, \mathcal{X})$  un sistema de coordenadas en torno de  $p \in M$  y  $\{X_i\}$  una base de  $T_pM$  en este sistema de coordenadas. Entonces:

$$K(X_i, X_j)X_k = \sum_{l=1}^n K_{ijk}^l X_l,$$

donde  $K_{ijk}^l$  son dadas por :

$$K_{ijk}^l = X_j \Gamma_{ik}^l - X_i \Gamma_{jk}^l + \sum_{s=1}^n \Gamma_{ik}^s \Gamma_{js}^l - \sum_{s=1}^n \Gamma_{jk}^s \Gamma_{is}^l.$$

**Observación 1.7.2** Si en las coordenadas  $(U, \mathcal{X})$  escribimos  $X = \sum_{i=1}^n u^i X_i$ ,  $Y = \sum_{j=1}^n v^j Y_j$ ,  $Z = \sum_{k=1}^n w^k Z_k$ , por la linealidad de  $K$  tenemos:

$$K(X, Y)Z = \sum_{i,j,k,l=1}^n K_{ijk}^l u^i v^j w^k X_l. \quad (1.13)$$

**Curvatura Seccional.** Íntimamente relacionado con el operador curvatura  $R$  está la curvatura seccional (o riemanniana) que definiremos a continuación

**Definición 1.7.2** Dado un espacio vectorial  $V$ , indicaremos por  $|x \wedge y|$  a la expresión

$$\sqrt{|x|^2|y|^2 - \langle x, y \rangle^2},$$

que representa el área del paralelogramo bidimensional definido por un par de vectores  $x, y \in V$

**Definición 1.7.3** Sea  $\sigma \subset T_p M$  un subespacio bidimensional del espacio vectorial  $T_p M$  y sean  $x, y \in \sigma$ , dos vectores linealmente independientes. Entonces,

$$K(x, y) = \frac{(K(x, y)x, y)}{|x \wedge y|},$$

no depende de la elección de los vectores  $x$  e  $y$ .

**Definición 1.7.4** (Curvatura Seccional). Dado un punto  $p \in M$  y un subespacio bidimensional  $\sigma \subset T_p M$ . El número  $K(x, y) = K(\sigma)$ , donde  $\{x, y\}$  es una base de  $\sigma$ , es llamado Curvatura Seccional de  $M$ .

Si  $K(x, y) \leq 0$  para todo  $x, y \in \sigma$  entonces, la curvatura seccional de la variedad

riemanniana es no positiva.

Si  $K(x, y) \geq 0$  para todo  $x, y \in \sigma$  entonces, la curvatura seccional de la variedad riemanniana es no negativa.

### Variedades completas

**Definición 1.7.5** Una variedad riemanniana  $M$  es llamada (geodésicamente) completa si para todo  $p \in M$ , la aplicación exponencial,  $\exp_p$ , están definidas para todos  $v \in T_p M$ ; es decir, las geodésicas que parten de  $p$  están definidas para todos los valores del parámetro  $t \in \mathbb{R}$ .

**Definición 1.7.6** Dados dos puntos  $p$  y  $q$  en  $M$ , la distancia riemanniana de  $p$  a  $q$  en la variedad, denotada por  $d(x, y)$ , es definida por

$$d(p, q) = \inf_{\gamma} \int_a^b \|\gamma'(t)\| dt \quad (1.14)$$

**Proposición 1.7.1** Con la distancia geodésica (1.14)  $M$  es un espacio métrico.

**Teorema 1.7.1** (Hopf-Rinow) Sea  $M$  una variedad riemanniana y sea  $p \in M$ .

Las siguientes afirmaciones son equivalentes:

- a)  $\exp_p$  está definida en todo  $T_p M$  para algún  $p \in M$
- b) Los limitados y cerrados de  $M$  son compactos.
- c)  $M$  es completo como espacios métrico.

d)  $M$  es geodésicamente completa.

e) Para todo  $q \in M$  existe una geodésica uniendo  $p$  y  $q$  con  $d(p, q) = \inf_{\gamma} \int_a^b \|\gamma'(t)\| dt$

**Teorema 1.7.2** En una variedad Riemanniana completa  $M$  de dimensión finita con curvatura seccional no-negativa, tenemos:

$$l_3^2 \leq l_1^2 + l_2^2 - 2l_1l_2 \cos \theta, \quad (1.15)$$

donde  $l_i$  denota la longitud de  $\alpha_i$ ,  $i = 1, 2$ ,  $l_3 = d(\alpha_1(l_1), \alpha_2(l_2))$  y

$\theta = \angle(\alpha_1'(0), \alpha_2'(0))$ .

**Definición 1.7.7** Sea  $M$  una variedad riemanniana y  $f : M \rightarrow \mathbb{R}$ , se define campo vectorial **grad(f)** (gradiente de la función  $f$ ) como el único campo vectorial que cumple:

$$df_p(X(p)) = \langle \text{grad}f(p), X(p) \rangle_x, \forall x \in \mathcal{H}. \quad (1.16)$$

**Observación 1.7.3** Sea  $M \subset \mathbb{R}^n$  una variedad riemanniana con la métrica definida por  $\langle v, w \rangle_x = v^T G(x)w$  donde  $G(x)$  es una matriz simétrica definida positiva.

Se puede caracterizar el campo gradiente como:

$$\text{grad}f(q) = G^{-1}(q)f'(q), \quad (1.17)$$

donde  $G^{-1} = g^{ij}(q)$  es la matriz inversa de  $G(q)$  y  $f'(x) = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})$  es el vector de derivadas parciales de la función  $f \circ X$ . En efecto,

$$\begin{aligned} df_p(X(p)) &= f'(q)^T v = f'(q)^T (G^{-1}(q))^T G(q) v \\ &= (G^{-1}(q) f'(q))^T G(q) v = \langle G^{-1}(q) f'(q), v \rangle_q. \end{aligned}$$

**Definición 1.7.8** Sea  $M$  una variedad Riemanniana y  $f : M \rightarrow \mathbb{R}$  una función de clase  $C^k$ ,  $k \geq 2$ . La Hessiana de  $f$ , denotada por  $H^f$  es definida como la derivada covariante del campo vectorial gradiente, esto es,

$$H^f = \frac{D}{dt}(\text{grad}f).$$

La Hessiana en el punto  $p$ , en la dirección de  $v \in T_p M$  es:

$$H_p^f(v, w) = \left\langle \frac{D_v}{dt}(\text{grad}f)(p), w \right\rangle_p. \quad (1.18)$$

**Proposición 1.7.3** Para todo  $X, Y \in \mathcal{H}$ , la Hessiana en términos de la conexión riemanniana vienen dada por:

$$\langle H_p^f X, Y \rangle = (XY - \nabla_X Y) f$$

**Observación 1.7.4** En términos de la base  $\{X_k\}$  tenemos

$$\begin{aligned} \langle H_p^f X_i, X_j \rangle &= (X_i X_j - \sum_{m=1}^n \Gamma_{ij}^m X_m) f \\ &= \left( \frac{\partial^2 f}{\partial x_i \partial x_j} - \sum_{m=1}^n \Gamma_{ij}^m \frac{\partial f}{\partial x_m} \right) f \end{aligned}$$

**Ejemplo 1.7.1** Sea la variedad riemanniana  $\mathbb{R}^n$  con métrica  $G(x) = I$ . Sabemos que  $\Gamma_{ij}^m = 0 \forall i, j, m = 1, 2, \dots, n$ , entonces la Hessiana riemanniana coincide con la Hessiana usual.

**Definición 1.7.9** Sea  $(X, d)$  un espacio métrico completo. Una sucesión  $\{y^k\}$ ,  $k \geq 0$ , de  $X$  es llamada *cuasi-Fejér convergente* a  $U \subset X$ , si para cada  $u \in U$  existe una sucesión  $\varepsilon_k \subseteq \mathbb{R}$  tal que,  $\varepsilon_k \geq 0$ ,  $\sum_{k=0}^{+\infty} \varepsilon_k < +\infty$  y

$$d^2(y^{k+1}, u) \leq d^2(y^k, u) + \varepsilon_k. \quad (1.19)$$

**Teorema 1.7.3** En un espacio métrico completo  $(X, d)$ . Si  $\{y^k\}$  es *cuasi-Fejér convergente* a un conjunto no vacío  $U \subset X$ , entonces  $\{y^k\}$  es acotada. Si, además, un punto de acumulación  $\bar{y}$  de  $\{y^k\}$  pertenece a  $U$ , entonces  $\{y^k\}$  converge y  $\lim_{k \rightarrow \infty} y^k = \bar{y}$ .

## 1.8. Preliminares de optimización

**Definición 1.8.1** Sea  $M$  una variedad Riemanniana y  $f : M \rightarrow \mathbb{R}$ , definimos por **Punto Crítico**, los puntos donde :

$$\langle \text{grad}f(x), v \rangle_x = 0, x \in M \text{ y } \forall v \in T_x(M). \quad (1.20)$$

**Definición 1.8.2** Sea  $f : M \rightarrow \mathbb{R}$ ,  $x^* \in M$  y una  $\mathbb{U}$  vecindad abierta de  $x^*$ , entonces,  $x^*$  es un **Punto Mínimo Local de  $f$** , si se cumple que  $f(x^*) \leq f(x)$   $\forall x \in \mathbb{U} \cap A$ ; si la desigualdad se cumple para todo  $x \in M$ , entonces  $x^*$  es un **Punto Mínimo Global de  $f$  en  $A$** .

**Definición 1.8.3** Se dice que una función  $f : M \rightarrow \mathbb{R}$  es *semicontinua inferiormente* en  $\hat{x} \in M$ , si para toda sucesión  $\{x^k\}$  de  $M$  convergente a  $\hat{x}$  se tiene



que:

$$\liminf_{k \rightarrow \infty} f(x^k) \geq f(\hat{x}).$$

Si  $f$  es semicontinua inferiormente para todo  $x \in M$ , entonces decimos que  $f$  es semicontinua inferiormente en  $M$ .

**Teorema 1.8.1** (Weierstrass) si  $f : M \rightarrow \mathbb{R}$  es semicontinua inferiormente y  $M$  es compacto, entonces existe un punto mínimo global de  $f$ .

**Teorema 1.8.2** (Condición necesaria de primer orden). Sea  $f : M \rightarrow \mathbb{R}$  de clase  $C^1$ . Si  $x^*$  es un punto de mínimo local, entonces  $\text{grad}f(x^*) = 0$ .

**Demostración 1.8.1** Sean  $v \in T_{x^*}M$  y  $\gamma : \mathbb{R} \rightarrow M$  una curva geodésica tal que  $\gamma(0) = x^*$  y  $\gamma'(0) = v$ . Definamos la aplicación  $h : \mathbb{R} \rightarrow \mathbb{R}$  tal que  $h(t) = f(\gamma(t))$ . Como  $x^*$  es punto de mínimo local de  $f$ , existe  $\varepsilon > 0$  tal que  $h(0) = f(x^*) \leq f(\gamma(t)) = h(t)$ , para todo  $t \in (-\varepsilon, \varepsilon)$ , por lo que en  $t = 0$  tenemos un punto de mínimo local de  $h$ . Por la condición necesaria de primer orden en  $\mathbb{R}$  se tiene  $h'(0) = \langle \text{grad}f(x^*), v \rangle = 0$ . En particular para  $v = \text{grad}f(x^*)$  tenemos que  $\text{grad}f(x^*) = 0$ .

**Teorema 1.8.3** (Condición necesaria de segundo orden). Sea  $f : M \rightarrow \mathbb{R}$  de clase  $C^2$ . Si  $x^*$  es un punto de mínimo local, entonces  $\langle v, H_{x^*}^f v \rangle \geq 0$ ,  $\forall v \in T_{x^*}M$ .

**Demostración 1.8.2** Sean  $v \in T_{x^*}M$  y  $\gamma : \mathbb{R} \rightarrow M$  una curva geodésica tal que  $\gamma(0) = x^*$  y  $\gamma'(0) = v$ . Definamos la aplicación  $h : \mathbb{R} \rightarrow \mathbb{R}$  tal que  $h(t) = f(\gamma(t))$ .

De la demostración del teorema anterior, sabemos  $t = 0$  hay un punto de mínimo local de  $h$ , entonces por la condición necesaria de segundo orden en  $\mathbb{R}$  se tiene,  $h'(0) = 0$ , luego  $h''(0) \geq 0$ . Ahora bien,

$$\begin{aligned} h'(t) &= \langle \text{grad}f(\gamma(t)), \gamma'(t) \rangle \\ h''(t) &= \left\langle \frac{D}{dt} \text{grad}f(\gamma(t)), \gamma'(t) \right\rangle + \left\langle \text{grad}f(\gamma(t)), \frac{D}{dt} \gamma'(t) \right\rangle \\ &= \langle H_{\gamma(t)}^f \gamma'(t), \gamma'(t) \rangle \\ &= \langle H_v^f v, v \rangle \geq 0. \end{aligned}$$

**Teorema 1.8.4** (Condición suficiente de segundo orden). Sea  $f : M \rightarrow \mathbb{R}$  de clase  $C^2$ . Si  $x^* \in M$  cumple:

- $\text{grad}f(x^*) = 0$ ,
- $H_{x^*}^f$  definida positiva,

entonces,  $x^*$  es un punto de mínimo local estricto de  $f$ .

**Definición 1.8.4** Sea  $A \subset M$  donde  $M$  es una variedad riemanniana completa.

Se dice que  $A$  es un **Conjunto Totalmente Convexo** si contiene toda geodésica  $\gamma$  de  $M$  que conecte a  $p$  y  $q$ , con  $p, q \in A$ .

**Definición 1.8.5** Sea  $A \subset M$  donde  $M$  es una variedad riemanniana completa.

Diremos que  $A$  es convexo si para todo par de puntos  $p$  y  $q$  de  $A$  existe una geodésica minimal que une  $p$  y  $q$  contenido en  $A$ .

**Definición 1.8.6** Sea  $f : M \rightarrow \mathbb{R}$ ; diremos que  $f$  es convexa si y solo si para toda geodésica  $\gamma : \mathbb{R} \rightarrow M$ , la compuesta  $f \circ \gamma : \mathbb{R} \rightarrow \mathbb{R}$  es convexa como una función real, es decir,  $\forall a, b \in \mathbb{R}, \forall \lambda \in [0, 1]$  se cumple:

$$f((\gamma(1 - \lambda)a + \lambda b)) \leq (1 - \lambda)f(\gamma(a)) + \lambda f(\gamma(b)). \quad (1.21)$$

**Teorema 1.8.5** Sea  $f : M \rightarrow \mathbb{R}$  y  $M$  una variedad Riemanniana. Se dice que  $f$  es convexa si y solo si para todo  $p \in M$  y para toda geodésica  $\gamma : [0, +\infty) \rightarrow M$  tal que  $\gamma(0) = p$  se cumple:

$$f(\gamma(t)) - f(p) \geq t \langle \text{grad}f(p), \gamma'(0) \rangle \text{ con } t \in [0, +\infty). \quad (1.22)$$

**Demostración 1.8.3** Supongamos que  $f$  es convexa, entonces para toda geodésica  $\gamma : \mathbb{R} \rightarrow M$ ;  $f \circ \gamma$  es convexa, en particular para  $h : [0, +\infty) \rightarrow \mathbb{R}$  dada por  $h(t) = f(\gamma(t))$  con  $t \in [0, +\infty)$ .

Como  $h$  es convexa como función real, entonces se cumple

$$h(t) \geq h(0) + th'(0). \quad (1.23)$$

Por otro lado

$$h'(t) = (f \circ \gamma)' = \langle \text{grad}f(\gamma(t)), \gamma'(t) \rangle,$$

evaluando  $h'$  en 0

$$\begin{aligned} h'(0) &= \langle \text{grad}f(\gamma(0)), \gamma'(0) \rangle \\ &= \langle \text{grad}f(p), \gamma'(0) \rangle \end{aligned}$$

sustituyendo la igualdad anterior en 1.23 tenemos:

$$h(t) \geq h(0) + t\langle \text{grad}f(p), \gamma'(0) \rangle$$

lo que implica que:

$$f(\gamma(t)) - f(p) \geq +t\langle \text{grad}f(p), \gamma'(0) \rangle.$$

Recíprocamente, supongamos que se cumple que:

$$f(\gamma(t)) - f(p) \geq t\langle \text{grad}f(p), \gamma'(0) \rangle \text{ con } t \in [0, +\infty),$$

con  $\gamma : \mathbb{R} \rightarrow M$  la geodésica tal que  $\gamma(0) = p$ .

Sea  $h : [0, +\infty) \rightarrow \mathbb{R}$  dada por  $h(t) = f(\gamma(t))$  con  $t \in [0, +\infty)$ .

como  $\gamma(0) = p$  y  $h'(0) = \langle \text{grad}f(p), \gamma'(0) \rangle$ , ocurre que  $h(t) - h(0) \geq th'(0)$ ,

de esto se tiene que  $h = f \circ \gamma$  es convexa y por lo tanto  $f$  es convexa.

**Corolario 1.8.1** Si  $f : M \rightarrow \mathbb{R}$  es convexa entonces todos los puntos críticos de  $f$  son mínimos globales de  $f$ .

**Demostración 1.8.4** Sea  $p \in M$  un punto crítico de  $f$ , entonces  $\text{grad}f(p) = 0$ .

Dado  $q \in M$  con  $q \neq p$ , existe por teorema de Hopf-Rinow's una geodésica  $\gamma : [a, b] \rightarrow M$  que conecta a  $p$  con  $q$  donde  $\gamma(a) = p$  y  $\gamma(b) = q$ . Como  $f$  es convexa por teorema de condición de primer orden para convexidad se cumple:

$f(\gamma(b)) - f(\gamma(a)) \geq \langle \text{grad}f(p), \gamma'(0) \rangle$  y dado que  $q$  es punto crítico  $f(q) \geq f(p)$   
 $\forall q$ . Por lo tanto, todo punto crítico de una función convexa es un mínimo global.

**Teorema 1.8.6** Sea  $M$  una variedad riemanniana, una función  $f : M \rightarrow \mathbb{R}$  es convexa si y solo si el Hessiano de  $f$  ( $H^f$ ) es semi-definido positivo.

**Demostración 1.8.5** Supongamos que  $f$  es convexa, sean  $p \in M$  y  $v \in T_p M$  y  $\gamma$  la única geodésica tal que  $\gamma(0) = p$  y  $\gamma'(0) = v$ . Definamos  $h : \mathbb{R} \rightarrow \mathbb{R}$  dada por  $f \circ \gamma = h$  por lo que  $h$  es convexa. Así, por la teoría de convexidad,  $h''(t) \geq 0 \forall t$ , ahora por definición de derivada

$$h'(t) = \langle \text{grad}f(\gamma(t)), v \rangle_{\gamma(t)},$$

luego

$$h''(0) = \langle H_p^f v, v \rangle \geq 0.$$

Recíprocamente, supongamos  $\langle H_p^f v, v \rangle \geq 0$ . Definiendo  $h : \mathbb{R} \rightarrow \mathbb{R}$  dada por  $f \circ \gamma = h$ , se tiene que  $h''(t) \geq 0$  y  $h = f \circ \gamma : \mathbb{R} \rightarrow \mathbb{R}$  es convexa, finalmente por la definición de convexidad  $f$  es convexa.

**Definición 1.8.7** Sea  $M$  una variedad riemanniana completa y  $f : M \rightarrow \mathbb{R}$  una función real.  $f$  es llamada cuasi-convexa en  $M$  si para todo  $x, y \in M$ ,  $t \in [0, 1]$ , se cumple:

$$f(\gamma(t)) \leq \max\{f(x), f(y)\},$$

para toda curva geodésica  $\gamma : [0, 1] \rightarrow M$ , tal que  $\gamma(0) = x$  y  $\gamma(1) = y$ .

**Teorema 1.8.7** Sea  $f : M \rightarrow \mathbb{R}$  una función diferenciable y cuasi-convexa en una variedad riemanniana completa  $M$  y sean  $x, y \in M$ . Si  $f(x) \leq f(y)$  entonces:

$$\langle \text{grad}f(y), \gamma'(0) \rangle \leq 0$$

donde  $\gamma$  es la curva geodésica tal que  $\gamma(0) = y$  y  $\gamma(1) = x$ .

**Definición 1.8.8** Sea  $M$  una variedad riemanniana, una función  $f : M \rightarrow \mathbb{R}$  es pseudoconvexa si, para todo par de puntos distintos  $x, y \in M$  y toda curva geodésica que une  $x$  a  $y$  ( $\gamma(0) = x$ ,  $\gamma(1) = y$ ) tenemos :

$$\langle \text{grad}f(x), \gamma'(0) \rangle \geq 0, \text{ entonces } f(y) \geq f(x).$$

**Definición 1.8.9** Sea  $\Gamma \geq 0$  una constante y  $f : M \rightarrow \mathbb{R}$ , si para cada  $p, q$  en  $M$  y el segmento geodésico  $\gamma : [0, a] \rightarrow M$  que conecta a  $p$  y  $q$  con  $\gamma(0) = p$  y  $\gamma'(a) = q$  dada por:

$l(\gamma_{0,t}) =$  longitud del segmento  $[\gamma(0), \gamma(t)] = \int_0^t \|\gamma'(t)\| dt$  cumple lo siguiente

$$\|\text{grad}f(\gamma(t)) - P_\gamma \text{grad}f(p)\| \leq \Gamma \|l(\gamma_{0,t})\|$$

se dice que  $f$  tiene gradiente  $\Gamma$ -lipschitziana.

**Definición 1.8.10** Sea  $f : M \rightarrow \mathbb{R}$ , una función da la variedad  $M$  a los reales, se denomina conjunto de subnivel  $a$ :  $M^a = \{x \in M : f(x) \leq a\}$  con  $a \in \mathbb{R}$

## 1.9. Búsqueda lineal no monótona

Presentamos algunas ideas principales acerca del uso de búsquedas lineales no-monótonas, esta constituye una técnica que ha sido usada con mucha frecuencia en los últimos años, en diferentes aplicaciones y extensiones de problemas de programación no lineal. Esta búsqueda surge como una ampliación de la conocida búsqueda de Armijo, [13], la cual fue la primera regla eficiente de búsquedas inexactas y que decrece el tamaño de las evaluaciones funcionales en cada iteración. Esta condición no es impuesta en las búsquedas no-monótonas, de esta forma permite que estas evaluaciones funcionales puedan aumentar y/o disminuir en cada iteración, de allí su nombre.

En cada iteración del método de búsqueda no-monótona, se calcula una dirección de descenso  $d^k$  y luego se decide cuanto moverse a lo largo de esta dirección. La iteración viene dada por:

$$x^{k+1} = x^k + \alpha_k d^k,$$

donde el tamaño de paso  $\alpha_k$  es calculado de la siguiente manera:

Dado  $a > 0$ ,  $\theta \in (0, 1)$  y  $\beta \in (0, 1)$ ,  $\alpha_k = \theta^{h_k} a$ , donde  $h_k$  es el primer entero no negativo  $h$  para el cual:

$$f(x^k + \alpha_k d^k) \leq \max_{0 \leq j \leq m(k)} [f(x^{k-j})] + \beta \theta^h a \nabla f(x^k)^T d^k,$$

donde  $m(0) = 0$  y  $0 \leq m_k \leq \min\{m(k-1) + 1, M\}$ ,  $k \geq 1$ .

La primera de estas estrategias fue propuesta por [11], quienes la usaron, junto con el método de Newton, en la minimización de una función dos veces continuamente diferenciable y hasta la fecha se han propuesto diferentes búsquedas no-monótonas, ver([12]).



---

## Capítulo 2

### Optimización en Variedades Riemannianas

---

En este capítulo estudiaremos algunos de los artículos que se han desarrollado en los últimos años, que tratan acerca de optimización en variedades Riemannianas. Se señalará que problema es propuesto en cada trabajo, los métodos plateados para resolverlo, así como también los resultados de convergencia, especificando en cada caso las hipótesis que fueron necesarias para ello. Se mostrarán de forma detallada algunas de las pruebas que servirán como justificación de los experimentos numéricos en el último capítulo de este trabajo.

#### 2.1. Método de Gradiente Proyectado a lo Largo de Geodésicas.

En el artículo [1], Luenberger considera el siguiente problema de optimización no lineal

$$\min_{s.a. h(x)=0} f(x) \quad (2.1)$$

donde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $m < n$ , son funciones de clase  $c^3$ .

Luenberger expone que cuando deseamos hallar el mínimo de una función  $f$  sobre una superficie  $S$ , definida implícitamente por  $S = \{x \in \mathbb{R}^n; h(x) = 0\}$ , la idea del método del gradiente proyectado consiste en moverse de un punto  $x^k$  en  $S$  a un nuevo punto  $x^{k+1}$  en  $S$ , en una dirección, determinada por la proyección del gradiente (negativo) de  $f$  en  $x^k$  sobre el plano tangente de  $S$  en  $x^k$ . El punto  $x^{k+1}$  es determinado como el mínimo en esa dirección. Para una Superficie no lineal  $S$ , la idea requiere una adicional definición, ya que el movimiento debe ser restringido a  $S$  y no simplemente a su plano tangente.

Para completar la idea del algoritmo hay que definir una curva  $x^k(\varepsilon)$ ,  $\varepsilon > 0$ , comenzando en  $x^k = x^k(0)$  y descansando en  $S$ . EL nuevo punto  $x^{k+1}$  es escogido como el primer mínimo local de  $f(x^k(\varepsilon))$ , para  $\varepsilon > 0$ .

En la practica, es difícil generar curvas en  $S$  y determinar con exactitud el punto minimizador a lo largo de la curva, por lo que usualmente nos movemos a lo largo del plano tangente en la dirección del gradiente proyectado a un nuevo punto favorable y luego trasladamos este punto a un punto cercano en  $S$ .

Existen dos procedimientos para generar las curvas  $x^k(\varepsilon)$ , ambos basados en pro-

yección. Sea  $y_k$  la proyección del gradiente negativo de  $f$  en  $x^k$  en el plano tangente en de  $S$  en  $x^k$ . Consideremos la línea  $x^k + \varepsilon y_k$ ,  $\varepsilon \geq 0$ . El primer método, para cada  $\varepsilon$ , sea  $x^k(\varepsilon)$  el punto en  $S$  más cercano a  $x^k + \varepsilon y_k$ . El punto  $x^k(\varepsilon)$  es encontrado por proyección perpendicular a  $S$ . El segundo método, para cada  $\varepsilon$ ,  $x^k(\varepsilon)$  es la proyección de  $x^k + \varepsilon y_k$  en  $S$  pero con la proyección perpendicular al plano tangente en  $x^k$  en lugar de perpendicular a  $S$ . Luenberger argumenta que el desarrollo del análisis de convergencia para el método del gradiente proyectado basado en los dos procedimientos descritos arriba, conduce a horrendas aproximaciones y a casi un sin fin de suposiciones, por lo cual propone que el movimiento sea a lo largo de geodésicas.

El método de Luenberger es como sigue:

**Algoritmo 2.1.1**

1. Dado  $x^k$ ,  $k \geq 1$ , un punto factible, calcular la proyección  $y_k$  de  $-\text{grad}f(x^k)$  sobre el plano tangente de la superficie  $(T_{x^k}S)$ , donde  $S = \{x \in \mathbb{R}^n; h(x) = 0\}$ .
2. Determinar la única geodésica  $\gamma(t)$  con  $t \geq 0$  de  $S$  tal que  $\gamma(0) = x^k$  y  $\gamma'(0) = y_k$
3. Minimizar  $f(\gamma(t))$ , obteniendo  $t_k$  (paso) y hacer  $x^{k+1} = \gamma(t_k)$ .

Luenberger prueba la convergencia global del algoritmo (2.1.1) a un punto crítico de  $f$  en  $S$  y estima la velocidad de convergencia a la vecindad de un punto crítico el cual es un mínimo local estricto.

## 2.2. Minimizando una Función Diferenciable Sobre una Variedad Diferenciable.

El problema estudiado por Gabay en el artículo [2] es el estudiando en (2.1) por Luenberger

$$\min_{s.a. h(x)=0} f(x) \tag{2.2}$$

donde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $m < n$ , son funciones de clase  $c^\sigma$ ,  $\sigma \geq 2$ , en el cual busca dar una generalización al método de máximo descenso.

Gabay comienza suponiendo que la siguiente hipótesis de regularidad se cumple: 0 es valor regular del mapa  $h$ , esto es, el Jacobiano  $h'(x)$  de  $h$  en  $x$  es de rango completo  $m$ , para todo  $x \in M = h^{-1}(0)$ . Bajo la suposición de regularidad, muestra que  $M = \{x \in \mathbb{R}^n; h(x) = 0\}$  es una variedad diferenciable, sobre la cual define su plano tangente  $(T_p M)$  y lo dota con una forma bilineal positiva  $\gamma_p : T_p M \times T_p M \rightarrow \mathbb{R}$ , llamada métrica riemanniana, la cual define una estructura riemanniana en  $M$ . Para entrar en el ambiente riemanniano, define los conceptos

de campo de vectores, diferenciación covariante, transporte paralelo y geodésicas, señalando dentro de este último la aplicación exponencial. Establece las condiciones de optimalidad en términos de la restricción de  $f$  a la variedad  $M$  y también define el campo gradiente de  $f$  en  $M$  y muestra que depende de la métrica riemanniana.

Gabay señala que para encontrar un mínimo local en  $\mathbb{R}^n$  de una función continuamente diferenciable  $f$ , el método de máximo descenso genera, a partir de una aproximación inicial  $x^0$ , aproximaciones sucesivas acordes con la siguiente iteración:

$$x^{k+1} = x^k + t_k p^k \quad k = 0, 1, 2, \dots, \quad (2.3)$$

donde  $p^k = -\nabla f(x^k)$ , es llamada dirección de máximo descenso, y  $t_k$  es llamado tamaño del paso, el cual es un escalar positivo seleccionado como el primer mínimo local en  $\mathbb{R}^+$  de la función  $g(t) = f(x^k + tp^k)$ ; se dice que  $t_k$  es seleccionado por una *búsqueda lineal exacta* y se denota la solución por:

$$t_k = \operatorname{argmin}\{f(x^k + tp^k) : t \geq 0\}. \quad (2.4)$$

Este método puede ser generalizado para encontrar un mínimo local de  $f$  sobre una variedad riemanniana  $M$ . La dirección de máximo descenso para  $f$  sobre  $M$

en  $x^k$  esta dada por:

$$p^k = -\nabla_M^\gamma f(x^k), \quad (2.5)$$

el opuesto del gradiente de  $f$  sobre la variedad riemanniana, donde  $\gamma$  es la métrica riemanniana.

Como las geodésicas en  $M$  juegan el rol de las líneas rectas en  $\mathbb{R}^n$ , Gabay define el *Método de descenso acelerado a lo largo de geodésicas* como la siguiente iteración:

$$x^{k+1} = \exp_{x^k}(t_k p^k) \quad k = 0, 1, 2, \dots, \quad (2.6)$$

donde  $p^k$  es definida por (2.5) y el tamaño del paso  $t_k$  es determinado por *búsqueda geodésica exacta*,

$$t_k = \operatorname{argmin}\{f(\exp_{x^k}(t p^k)) : t \in t \geq 0\}. \quad (2.7)$$

El algoritmo dado por (2.5), (2.6), (2.7), fue analizado por Luenberger, quien explícitamente uso la estructura riemanniana  $\gamma^E$  en  $M$ , la inducida por la estructura euclidiana en  $\mathbb{R}^n$ . Gabay introduce algunas variantes al método propuesto por Luenberger en un marco totalmente intrínseco, es decir, la dirección de descenso que considerada es la determinada por el gradiente de la función objetivo relativo a la métrica riemanniana definida sobre la variedad  $M$ ; él prueba que la sucesión  $\{x^k\}$  generada por método de descenso acelerado a lo largo de geodésicas (2.5), (2.6), (2.7), está bien definida; o bien es finita, terminando en un punto crítico, o es infinita y existe una subsucesión que converge a un punto crítico.

## 2.3. Algoritmos Geodésicos en Geometría Riemanniana.

En el artículo [3] Neto, Lima y Oliveira, consideran el problema (2.1) estudiado por Luenberger.

$$\min_{s.a. h(x)=0} f(x) \quad (2.8)$$

donde  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ ,  $m < n$ , son funciones de clase  $c^\sigma$ ,  $\sigma \geq 2$ .

Nuevamente se considera la variedad  $M = \{x \in \mathbb{R}^n; h(x) = 0\}$ , dotada con una métrica riemanniana y se definen diversos conceptos tales como : campos de vectores, campo gradiente, Hessiano, transporte paralelo, derivada covariante, geodésicas, variedad completa, teorema de Hopf-Rinow's, tensor curvatura y ley de los cosenos. También se establecen las condiciones de optimalidad.

El algoritmo propuesto aquí, tiene sus dos primeros pasos iguales a los del algoritmo de Luenberger:

1. Dado  $x^k$ ,  $k \geq 1$  factible, calcular la dirección de descenso  $y_k$ , dada por:

$$y_k = -grad f(x^k).$$

2. Determinar la única geodésica  $\gamma(t)$ ,  $t \geq 0$  de M tal que  $\gamma(0) = x^k$  y  $\gamma'(0) =$

$$y_k.$$

El tercer paso en el algoritmo de Luenberger es reemplazado por los siguientes algoritmos intrínsecos:

**Algoritmo 2.3.1 (Paso Fijo)**

Dados  $\delta_1 > 0$   $\delta_2 > 0$  tal que  $\delta_1\Gamma + \delta_2 < 1$ , donde  $\Gamma$  es la constante de Lipschitz asociada al gradiente de  $f$ ; elegir

$$t_k \in \left(\delta_1, \frac{2(1 - \delta_2)}{\Gamma}\right).$$

**Algoritmo 2.3.2 (Búsqueda de Armijo)**

Elegir  $t_k = 2^{-i_k\bar{t}}$ , donde  $\bar{t} > 0$  es dado;  $i_k$  es el menor número natural positivo tal que

$$f(\gamma(t_k)) \leq f(x^k) - \beta t_k^2 \|\text{grad } f(x^k)\|^2, \text{ con } \beta \in (0, 1). \quad (2.9)$$

Neto, Lima y Oliveira prueban que si la función  $f : M \rightarrow \mathbb{R}$  es convexa y  $M$  es una variedad Riemanniana con curvatura no-negativa, completa, la sucesión  $\{x^k\}$  es generada por los algoritmos (2.3.1) y (2.3.2) converge globalmente a un punto mínimo de  $f$ ; extendiendo así la metodología usada por Luenberger a un marco más general de variedades riemannianas con curvatura no-negativa.



## 2.4. Una Regularización proximal del Método de descenso en Variedades Riemannianas.

El problema estudiado por Neto, Ferreira y Pérez en artículo [4] es el siguiente:

$$\min_{s.a. x \in M} f(x),$$

donde  $M$  es una variedad riemanniana completa y  $f : M \rightarrow \mathbb{R}$  es una función continuamente diferenciable. Es introducida una regularización proximal en la búsqueda lineal del tamaño del paso  $t_k$ . El método de máximo descenso clásico siguiente:

### Algoritmo 2.4.1

Tomar  $x^0 \in M$ .

Para  $k = 0, 1, 2, \dots$ ,

$$\begin{aligned} t_k &= \operatorname{argmin}_{t \geq 0} f(\exp_{x^k}(-t \operatorname{grad} f(x^k))) \\ x^{k+1} &= \exp_{x^k}(-t_k \operatorname{grad} f(x^k)), \end{aligned}$$

es reemplazado por el siguiente algoritmo:

**Algoritmo 2.4.2**

*Este algoritmo requiere una sucesión exógena,  $\{\lambda_k\}$ , de números reales, tal que, para todo  $k$ ,  $\lambda' \leq \lambda_k \leq \lambda''$ , donde  $0 < \lambda' \leq \lambda''$ .*

*Tomar  $x^0 \in M$ .*

*Para  $k = 0, 1, 2, \dots$ ,*

$$\varphi_k(t) = f(\exp_{x^k}(-t \operatorname{grad} f(x^k))) + t^2 \lambda_k \|\operatorname{grad} f(x^k)\|^2,$$

*calcular*

$$t_k = \operatorname{argmin}_{t \geq 0} \varphi_k(t),$$

*y hacer*

$$x^{k+1} = \exp_{x^k}(-t_k \operatorname{grad} f(x^k)).$$

Neto, Ferreira y Pérez muestran que la sucesión generada por el algoritmo (2.4.2), está bien definida y prueban dos tipos de convergencia.

**Convergencia débil**

Sin hipótesis concernientes a la curvatura de la variedad riemanniana  $M$  y convexidad de  $f$ , probaron que si el conjunto de nivel de  $f$  en  $x^0$  es acotado por debajo, la sucesión generada por el algoritmo (2.4.2), converge débilmente, a saber, es acotada, la distancia entre los iterados consecutivos tiende a cero y todos sus puntos de acumulación son puntos críticos.

**Convergencia plena**

La convergencia plena se cumple bajo la suposición de convexidad y de curvatura

seccional no-negativa de  $M$ , pero no se necesita la existencia conjuntos de nivel de  $f$  acotados. Considerando esto, los autores probaron que la sucesión generada por el algoritmo (2.4.2), converge a un punto minimizador de  $f$ .

## 2.5. Método de Descenso Para Minimización Cuasiconvexa en Variedades Riemannianas.

En el artículo [5] y [9] se extiende la convergencia plena del algoritmo de máximo descenso con una búsqueda de Armijo generalizada y una regularización proximal para resolver un problema de minimización cuasiconvexa en una variedad riemanniana completa.

El problema tratado es el siguiente:

$$\min_{x \in M} f(x), \quad (2.10)$$

donde  $M$  es una variedad riemanniana completa de dimensión finita y  $f : M \rightarrow \mathbb{R}$  es una función cuasiconvexa continuamente diferenciable en  $M$ .

El método de máximo descenso genera una sucesión  $\{x^k\}$  dada por :

$$x^0 \in M \quad (2.11)$$

$$x^{k+1} = \exp_{x^k}(-t_k \text{grad}f(x^k)) \quad (2.12)$$

donde  $\exp$  es la aplicación exponencial y  $t_k$  es un tamaño del paso.

Quiroz, Quispe y Oliveira comienzan suponiendo lo siguiente:

**Hipótesis A1**

El conjunto de mínimos globales del problema (2.10), denotado por  $X^*$  es no vacío, además, denotan por  $f^*$  el valor óptimo de ese problema y definen el siguiente conjunto no vacío  $U := \{x \in M : f(x) < \inf_k f(x^k)\}$ .

El siguiente lema es utilizado para probar que la sucesión, generada el método de máximo descenso es cuasi-Fejér convergente a  $U$ .

**Lema 2.5.1** *Sea  $f : M \rightarrow \mathbb{R}$  una función continuamente diferenciable y cuasi-convexa en una variedad riemanniana conexa, completa y de dimensión finita con curvatura seccional no negativa, entonces*

$$d^2(x^{k+1}, x) \leq d^2(x^k, x) + t_k^2 \| \text{grad}f(x^k) \|^2,$$

para todo  $x \in U$  y todo  $t_k > 0$ .

**Demostración 2.5.1** *Sea  $x \in U$  arbitrario. Supongamos que  $\gamma_1 : [0, 1] \rightarrow M$  es la geodésica minimal que une  $x^k$  y  $x$ , tal que  $\gamma_1(0) = x^k$ ,  $\| \gamma_1'(0) \| = 1$  y  $\gamma_2 : [0, 1] \rightarrow M$  es la geodésica minimal que une  $x^k$  y  $x^{k+1}$ , con  $\gamma_2'(0) = -t_k \text{grad}f(x^k)$ . Por propiedad de homogeneidad de las geodésicas,  $\gamma_2$  puede ser reparametrizada de la siguiente manera:*

$$\gamma_2 : [0, t_k \| \text{grad}f(x^k) \|] \rightarrow, \text{ tal que, } \gamma_2(t_k \| \text{grad}f(x^k) \|) = x^{k+1},$$

y así tenemos

$$\| \gamma_2'(0) \| = 1.$$

Por la ley de los cosenos tenemos:

$$d^2(x^{k+1}, x) \leq d^2(x^k, x) + t_k^2 \| \text{grad}f(x^k) \|^2 + 2t_k d(x^k, x) \langle \text{grad}f(x^k), \gamma'(0) \rangle.$$

Por otro lado como  $f$  es cuasiconvexa y  $f(x) \leq f(x^k)$  por A1, usando el teorema 1.8.7 se tiene que:  $\langle \text{grad}f(x^k), \gamma'(0) \rangle \leq 0$ , por lo que

$$d^2(x^{k+1}, x) \leq d^2(x^k, x) + t_k^2 \| \text{grad}f(x^k) \|^2.$$

El método de máximo descenso con una búsqueda de Armijo generalizada propuesto es el siguiente:

$$x^0 \in M \tag{2.13}$$

$$x^{k+1} = \text{exp}_{x^k}(-t_k \text{grad}f(x^k)) \tag{2.14}$$

donde el tamaño del paso  $t_k$  es determinado por:

$$t_k = \text{argmax}\{t : f(\text{exp}_{x^k}(-t_k \text{grad}f(x^k))) \leq f(x^k) - \alpha t \|\text{grad}f(x^k)\|^2, t = 2^{-i}, i = 0, 1, \dots\}, \text{ con } \alpha \in (0, 1)$$

En la primera parte de este artículo, el propósito es probar la convergencia plena de este método para el caso cuasiconvexo; y para ello, fueron necesarias las siguientes suposiciones :

**Hipótesis A2** Sea  $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  una función tal que :

**A2.1** Existe  $\alpha \in (0, 1)$ ,  $\tau_\alpha > 0$ , tal que  $\forall t \in (0, \tau_\alpha] : \phi(t) \leq \alpha t$ ,

**A2.2** Existe  $\beta > 0$ ,  $\tau_\beta \in (0, +\infty)$ , tal que  $\forall t \in (0, \tau_\beta) \cap \mathbb{R} : \phi(t) \geq \beta t^2$ ,

**A2.3** Para todo  $k$ ,  $f(x^{k+1}) \leq f(x^k) - \phi(t_k) \|\text{grad}f(x^k)\|^2$  y  $0 < t_k \leq \tau_\beta$ , en (2.14)

**A2.4** Existe  $\gamma > 1$ ,  $\tau_\gamma > 0$  tal que  $\forall k : t_k \geq \tau_\gamma$  ó

Existe  $\bar{t}_k \in [t_k, \gamma t_k] : f(\exp_{x^k}(-\bar{t}_k \text{grad}f(x^k))) \geq f(x^k) - \phi(\bar{t}_k) \|\text{grad}f(x^k)\|^2$ .

La proposición que viene a continuación muestra que la función es cuasi-Fejér convergente a  $U$ .

**Proposición 2.5.1** *Sea  $f : M \rightarrow \mathbb{R}$  una función continuamente diferenciable y cuasiconvexa. Suponga que las hipótesis A1 y A2 son satisfechas. Entonces la sucesión  $\{x^k\}$  generada por el método del gradiente con regla de Armijo generalizada es cuasi-Fejér convergente a  $U$ .*

**Demostración 2.5.2** *De las hipótesis A2.2 y A2.3 tenemos que :*

$$\beta t_k^2 \|\text{grad}f(x^k)\|^2 \leq f(x^k) - f(x^{k+1}), \quad (2.15)$$

lo cual implica

$$\sum_{k=0}^{+\infty} t_k^2 \|\text{grad}f(x^k)\|^2 \leq \frac{f(x^0) - f(x^*)}{\beta} < +\infty,$$

luego del Lemma (2.5.1) tenemos que la sucesión  $\{x^k\}$  generada por el método del gradiente con regla de Armijo generalizada es cuasi-Fejér convergente a  $U$ .

El siguiente teorema garantiza la convergencia de la sucesión

**Teorema 2.5.1** *Sea  $f : M \rightarrow \mathbb{R}$  una función continuamente diferenciable y cuasiconvexa. Suponga que las hipótesis A1 y A2 son satisfechas. Entonces la suce-*

ción  $\{x^k\}$  generada por el método del gradiente con regla de Armijo generalizado converge. Además, converge para un punto estacionario (un punto  $\bar{x}$  tal que  $\text{grad}f(\bar{x}) = 0$ ).

**Demostración 2.5.3** De la Proposición anterior, tenemos que  $\{x^k\}$  es cuasi-Fejér convergente a  $U$ , así por el teorema (1.7.3),  $\{x^k\}$  es acotada. Entonces existen  $\bar{x}$  y una subsucesión  $\{x^{k_j}\}$  de  $\{x^k\}$  que converge para  $\bar{x}$ . Por otro lado, de la continuidad de  $f$  tenemos que:

$$\lim_{j \rightarrow +\infty} f(x^{k_j}) = f(\bar{x}).$$

De (2.15), tenemos que  $\{f(x^k)\}$  es una sucesión no creciente, con una subsucesión que converge para  $f(\bar{x})$ , toda la sucesión converge para  $f(\bar{x})$  y por lo tanto

$$f(\bar{x}) \leq f(x^k), \quad \forall k \in \mathbb{N},$$

lo que implica que  $\bar{x} \in U$ . Nuevamente utilizando el teorema (1.7.3), concluimos que  $\{x^k\}$  converge para  $\bar{x}$ .

Ahora que queremos probar que  $\text{grad}f(\bar{x}) = 0$ . Por reducción al absurdo que  $\text{grad}f(\bar{x}) \neq 0$ . Es claro que  $\text{grad}f(x^k) \rightarrow \text{grad}f(\bar{x}) \neq 0$  y  $f(x^k) \rightarrow f(\bar{x})$ , luego de (2.15), tenemos que, usando el lema del emparedado

$$\lim_{k \rightarrow +\infty} t_k = 0. \tag{2.16}$$

Por otro lado usando A2.4 y A2.1 tenemos, para  $k$  suficientemente grande:

$$f(\exp_{x^k}(-\bar{t}_k \operatorname{grad} f(x^k))) - f(x^k) \geq -\alpha \bar{t}_k \|\operatorname{grad} f(x^k)\|^2.$$

Además, del teorema del valor medio, para cada  $k$ , existe  $t_k^* \in [0, \bar{t}_k]$  tal que :

$$-\langle \operatorname{grad} f(\exp_{x^k}(-\bar{t}_k \operatorname{grad} f(x^k))), P_{\gamma_k, 0, t_k^*} \operatorname{grad} f(x^k) \rangle \geq -\alpha \|\operatorname{grad} f(x^k)\|^2, \quad (2.17)$$

donde  $P_{\gamma_k, 0, t_k^*}$  es el transporte paralelo a lo largo de la geodésica  $\gamma_k$  tal que  $\gamma(0) = x^k$  y  $\gamma'(0) = -\operatorname{grad} f(x^k)$ . Ahora bien, de la ecuación (2.16) y la hipótesis A2.4, tenemos que  $\lim_{k \rightarrow +\infty} t_k^* = 0$ . Haciendo  $k \rightarrow +\infty$  en (2.17) y considerando la continuidad de  $\operatorname{grad} f$ ,  $\exp$  y el transporte paralelo, tenemos que  $1 \leq \alpha$ , lo cual contradice A2.1. Por tanto,  $\operatorname{grad} f(\bar{x}) = 0$ .

Una consecuencia del teorema anterior, es el siguiente resultado

**Corolario 2.5.1** *Sea  $f : M \rightarrow \mathbb{R}$  una función continuamente diferenciable y pseudoconvexa. Entonces, con las hipótesis A1 y A2, la sucesión  $\{x^k\}$  converge para un punto de mínimo global del problema planteado.*

En la segunda parte del artículo, se presenta el método de descenso con regularización proximal. Sea  $\{\lambda_k\}$  una sucesión de números reales tal que  $\lambda' \leq \lambda_k \leq \lambda''$ , donde  $0 < \lambda' \leq \lambda''$ . El método de máximo descenso con una regularización proximal genera una sucesión  $\{x^k\}$  definida por (2.11) y (2.12), donde

$$t_k = \operatorname{argmin} f \{ (\exp_{x^k}(-t \operatorname{grad} f(x^k))) + t^2 \lambda_k \|\operatorname{grad} f(x^k)\|^2 : t \geq 0 \}. \quad (2.18)$$



Como ya hemos mencionado Neto, Ferreira y Pérez probaron la convergencia de este método para el caso convexo, Quiroz, Quispe y Oliveira en este artículo, demuestran que si  $f : M \rightarrow \mathbb{R}$  es una función continuamente diferenciable y cuasiconvexa y además la hipótesis A1 es satisfecha. Entonces, la sucesión  $\{x^k\}$ , generada por (2.11), (2.12) y (2.18), converge a un punto estacionario.

Por último, se presentan experimentos numéricos para resolver problemas de optimización para funciones cuasiconvexas en la variedad  $\mathbb{R}_{++}^n$  y en el hipercubo abierto, usando la métrica diagonal.

## 2.6. Un algoritmo de descenso para optimización cuasiconvexa en variedades Riemannias

En este artículo [6], se analiza un algoritmo basado en direcciones de descenso para resolver el problema de minimizar una función cuasiconvexa sobre una variedad de Riemann. El tamaño de paso se obtiene mediante la regla de Armijo.

El problema tratado es el siguiente:

$$\min_{x \in M} f(x), \quad (2.19)$$

donde  $M$  es una variedad Riemanniana completa de dimensión finita y  $f : M \rightarrow \mathbb{R}$  es una función cuasiconvexa continuamente diferenciable en  $M$ .

Para espacios Euclidianos, se dice que  $d_x$  es una dirección suficiente de descenso si existen dos constantes positivas  $\gamma_0$  y  $\gamma_1$ , independientes de  $x$ , tal que :

$$d_x^T \nabla f(x) \leq -\gamma_0 \|\nabla f(x)\|_2^2 \quad (2.20)$$

$$\|d_x\|_2 \leq \gamma_1 \|\nabla f(x)\|_2, \quad (2.21)$$

donde  $\|\cdot\|_2$  es la norma euclidiana. Se extiende la definición de dirección suficiente de descenso para optimización en variedades riemannianas (DSDR) como sigue:

Consideremos números reales  $\alpha$  y  $\beta$  tal que  $d^k \in T_{x^k} M$  satisfice:

$$\langle \text{grad}f(x^k), d^k \rangle \leq -\alpha \|\text{grad}f(x^k)\|^2 \quad (2.22)$$

$$\|d^k\| \leq \beta \|\text{grad}f(x^k)\|, \quad (2.23)$$

donde el producto y la norma son dados por la métrica riemanniana.

Antes de describir el método se denota  $L_\lambda = \{x \in M : f(x) \leq \lambda\}$  y en adición, dado un punto  $x^k$  se define el siguiente conjunto:

$$D_k = \{d \in T_{x^k} M : -d \in N(x^k, L_f(x^k)), d \text{ es DSDR}\}, \quad (2.24)$$

donde  $N(x^k, L_f(x^k))$  es el cono normal de  $L_f(x^k)$  en el punto  $x^k$ .

Se presenta el siguiente algoritmo:

**Algoritmo 2.6.1 (Algoritmo suficiente de descenso)**

Sean  $x^0 \in M$ ,  $\theta \in (0, 1)$  y  $\tau \in (0, 1/2)$ .

Dado  $x^k \in M$  y  $d^k \in D_k$ ,

si  $\text{grad}f(x^k) = 0$  parar

si no, encontrar

$$x^{k+1} = \exp_{x^k}(\theta_k d^k), \quad (2.25)$$

donde

$$\theta_k = \text{argmax}\{\theta^j : f(\exp_{x^k}(\theta^j d^k)) - f(x^k) \leq \tau \theta^j \langle \text{grad}f(x^k), d^k \rangle, j \in \mathbb{N}\}. \quad (2.26)$$

Se muestra que el conjunto  $D_k$  es no vacío y por tanto la sucesión generada por (2.25) siempre existe debido a la existencia de  $\theta_k$ . A partir de aquí se considera que el algoritmo (2.6.1) genera una sucesión infinita  $\{x^k\}$ .

Teniendo en cuenta lo anterior, Sousa, Santos y Neto demuestran lo siguiente:

- a) La sucesión  $\{f(x^k)\}$  es noincreciente.
- b) La sucesión  $\{x^k\}$  es cuasi-Féjer convergente a  $\mathbb{U}$ .
- c) La sucesión  $\{x^k\}$  converge a un punto estacionario.

Se presentan experimentos numéricos para resolver problemas de optimización para funciones cusiconvexas en la variedad  $\mathbb{R}_{++}^n$  y en el hipercubo abierto, usando la métrica diagonal, los cuales son comparados con Papa Quiroz en [9].

---

## Capítulo 3

### Características de las métricas diagonales

---

En esta sección se derivan algunas propiedades geométricas de la métrica riemanniana diagonal

$$G(x) = \text{diag}\left(\frac{1}{h_1(x_1)^2}, \frac{1}{h_2(x_2)^2}, \dots, \frac{1}{h_n(x_n)^2}\right),$$

definida en el ortante positivo  $\mathbb{R}_{++}^n$ , cuya matriz inversa  $G^{-1}$  es:

$$G^{-1}(x) = \text{diag}(h_1(x_1)^2, h_2(x_2)^2, \dots, h_n(x_n)^2),$$

donde las funciones  $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$  son diferenciables. Se obtendrán expresiones simples para los símbolos de Christoffel, las ecuaciones explícitas para obtener las curvas geodésicas. Se prueba la curvatura nula de la variedad riemanniana  $(\mathbb{R}_{++}^n, G(x))$  y se dan expresiones explícitas para el gradiente y Hessiano. Estas expresiones nos serán de gran utilidad en la última parte de este trabajo, donde presentaremos algunos experimentos numéricos de optimización en variedades riemannianas.

1. Símbolos de Christoffel.

La relación entre métrica y Símbolos de Christoffel está dada por la ecuación (1.12), en este caso cuando  $k \neq m$  tenemos que  $g^{mk} = 0$  y así la expresión se reduce a:

$$\Gamma_{ij}^m = \frac{1}{2} \left\{ \frac{\partial}{\partial x_i} g_{jm} + \frac{\partial}{\partial x_j} g_{mi} - \frac{\partial}{\partial x_k} g_{ij} \right\} g^{mm}.$$

Se presentan dos casos:

Caso  $i = j$

$$\Gamma_{ii}^m = \frac{1}{2} \left\{ \frac{\partial}{\partial x_i} g_{im} + \frac{\partial}{\partial x_i} g_{mi} - \frac{\partial}{\partial x_k} g_{ii} \right\} g^{mm}.$$

Si  $m = i$

$$\Gamma_{ii}^i = -\frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i}.$$

si  $i \neq m$

$$\Gamma_{ii}^m = 0.$$

Caso  $i \neq j$

$$\Gamma_{ij}^m = \frac{1}{2} \left\{ \frac{\partial}{\partial x_i} g_{jm} + \frac{\partial}{\partial x_j} g_{mi} \right\} g^{mm}.$$

Si  $m = i$ , entonces  $m \neq j$  y

$$\Gamma_{ij}^i = 0.$$

Si  $m = j$ , entonces  $m \neq i$  y

$$\Gamma_{ij}^j = 0.$$

Si  $m \neq i$  y  $m \neq j$ , entonces

$$\Gamma_{ij}^m = 0.$$

En resumen tenemos:

$$\Gamma_{ij}^m = -\frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \delta_{im} \delta_{ij}, \quad (3.1)$$

la expresión de los Símbolos de Christoffel en relación a la métrica  $G(x)$ .

### Ejemplo 3.0.1

- Si  $h_i(x_i) = 1$ , entonces,  $G(x) = I$  y  $\Gamma_{ij}^m = 0$ ,  $\forall i, j, m = 1, \dots, n$ .
- Si  $h_i(x_i) = x_i$ , entonces,  $G(x) = X^{-2}$  y  $\Gamma_{ij}^m = -\frac{1}{x_i} \delta_{im} \delta_{ij}$
- Si  $h_i(x_i) = x_i^{\frac{r}{2}}$ , entonces,  $G(x)X^{-r}$  y  $\Gamma_{ij}^m = \frac{-r}{2} \frac{1}{x_i} \delta_{im} \delta_{ij}$

## 2. Derivada covariante y transporte paralelo.

Dado un campo de vectores  $V = \sum_{i=1}^n v_i X_i$  en la curva  $x(t)$ ,

$x(t) = (x_1(t), x_2(t), \dots, x_n(t)) \in \mathbb{R}_{++}^n$ , con  $x(0) = p = (p_1, p_2, \dots, p_n)$ , donde

$\{X_i\}_{i=1,2,\dots,n}$  es una base para  $T_p(\mathbb{R}_{++}^n) = \mathbb{R}^n$ , de (1.7) tenemos :

$$\frac{DV}{dt} = \sum_{i=1}^n \left( \frac{dv_i}{dt} + \sum_{j=1}^n v_j \frac{dx_j}{dt} \Gamma_{ij}^k \right) X_k, \quad (3.2)$$

sustituyendo los Símbolos de Christoffel (3.1) en la ecuación anterior obtenemos:

$$\frac{DV}{dt} = \sum_{i=1}^n \left( \frac{dv_i}{dt} - \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} v_i \frac{dx_i}{dt} \right) X_i. \quad (3.3)$$

Ahora bien, para  $V_0 = (V_1^0, V_2^0, \dots, V_n^0) \in T_p(\mathbb{R}_{++}^n)$ , la ecuación diferencial que define el transporte paralelo  $V(t) = (V_1(t), V_2(t), \dots, V_n(t))$  a lo largo de la curva  $x(t)$  es:

$$\frac{dV_i}{dt} - \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} V_i \frac{dx_i}{dt} = 0, \quad \forall i = 1, \dots, n \quad (3.4)$$

con la condición

$$V(0) = V_0.$$

**Afirmación 3.0.1** *La solución de la ecuación diferencial dada en (3.4) esta dada por:*

$$V_i(t) = v_i^0 \frac{h_i(x_i(t))}{h_i(p_i)}, \quad \forall i = 1, \dots, n \quad (3.5)$$

En efecto,

$$\frac{dV_i}{dt} = \frac{v_i^0}{h_i(p_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \frac{dx_i}{dt},$$

sustituyendo en la ecuación (3.4) obtenemos:

$$\begin{aligned} & \frac{v_i^0}{h_i(p_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \frac{dx_i}{dt} - \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \left( v_i^0 \frac{h_i(x_i(t))}{h_i(p_i)} \right) \frac{dx_i}{dt} \\ &= \frac{v_i^0}{h_i(p_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \frac{dx_i}{dt} - \frac{v_i^0}{h_i(p_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \frac{dx_i}{dt} = 0, \end{aligned}$$

además

$$V_i(0) = v_i^0 \frac{h_i(x_i(0))}{h_i(p_i)} = v_i^0 \frac{h_i(p_i)}{h_i(p_i)} = v_i$$

y así  $V(0) = V_0$ . De aquí, podemos definir el transporte paralelo a lo largo de la curva  $x(t)$  como la aplicación  $P_{x(t)} : T_p(\mathbb{R}_{++}^n) \rightarrow T_{x(t)}(\mathbb{R}_{++}^n)$  tal que

$$P_{x(t)}(V) = \left( v_1 \frac{h_1(x_1(t))}{h_1(p_1)}, v_2 \frac{h_2(x_2(t))}{h_2(p_2)}, \dots, v_n \frac{h_n(x_n(t))}{h_n(p_n)} \right). \quad (3.6)$$

### Ejemplo 3.0.2

- Si  $h_i(x_i) = x_i^{\frac{r}{2}}$ , entonces,  $P_{x(t)}(V)_i = v_i x_i^{r/2} / p_i^{r/2}$ .
- Si  $h_i(x_i) = x_i^{\frac{r}{2}}(1-x_i)^{\frac{r}{2}}$ , entonces,  $P_{x(t)}(V)_i = v_i x_i^{\frac{r}{2}}(1-x_i)^{\frac{r}{2}} / p_i^{\frac{r}{2}}(1-p_i)^{\frac{r}{2}}$ .

### 3. Ecuación Geodésica.

Sea  $p = (p_1, p_2, \dots, p_n) \in \mathbb{R}_{++}^n$ ,  $v = (v_1, v_2, \dots, v_n) \in T_p \mathbb{R}_{++}^n$  y  $x : I \rightarrow \mathbb{R}_{++}^n$ ,  $x(t) = (x_1(t), x_2(t), \dots, x_n(t))$  una curva tal que  $x(0) = p$  y  $dx(0)/dt = v$ .

Sustituyendo los símbolos de Christoffel en la ecuación (1.10)

$$\frac{d^2 x_i}{dt^2} - \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \left( \frac{dx_i}{dt} \right)^2 = 0, \quad i = 1, 2, \dots, n, \quad (3.7)$$

con las condiciones iniciales:

$$x_i(0) = p_i, \quad i = 1, 2, \dots, n,$$

$$x'_i(0) = v_i, \quad i = 1, 2, \dots, n.$$

La ecuación diferencial (3.7) es equivalente a resolver:

$$\frac{dx_i}{dt} = h_i(x_i) a_i, \quad (3.8)$$



para alguna constante  $a_i$ .

En efecto,

$$\frac{d^2x_i}{dt^2} = a_i \frac{\partial(h_i(x_i))}{\partial x_i} \frac{dx_i}{dt} \quad y \quad \left(\frac{dx_i}{dt}\right)^2 = h_i^2(x_i)a_i^2,$$

sustituyendo en la ecuación (3.7) obtenemos

$$\begin{aligned} \frac{d^2x_i}{dt^2} - \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \left(\frac{dx_i}{dt}\right)^2 &= a_i \frac{\partial(h_i(x_i))}{\partial x_i} \frac{dx_i}{dt} - \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} h_i^2(x_i)a_i^2 \\ &= a_i \frac{\partial(h_i(x_i))}{\partial x_i} \left(\frac{dx_i}{dt} - h_i(x_i)a_i\right) = 0. \end{aligned}$$

Por otro lado, la ecuación (3.8) es equivalente a resolver:

$$\int \frac{1}{h_i(x_i)} dx_i = a_i t + b_i \quad i = 1, 2, \dots, n \quad (3.9)$$

para algunas constantes  $a_i, b_i \in \mathbb{R}$

Entonces, la única geodésica  $x(t)$  de  $\mathbb{R}_{++}^n$  con métrica  $G(x)$ , pasando por el punto  $x(0) = p$ , en la dirección  $x'(0) = v$ , es obtenida resolviendo el siguiente problema:

$$\int \frac{1}{h_i(x_i)} dx_i = a_i t + b_i \quad i = 1, 2, \dots, n \quad (3.10)$$

donde  $a_i, b_i$  son constantes reales tales que :

$$x_i(0) = p_i, \quad i = 1, 2, \dots, n,$$

$$x'_i(0) = v_i, \quad i = 1, 2, \dots, n.$$

### Ejemplo 3.0.3 *Ejemplos de curvas Geodésicas*

- Si  $h_i(x_i) = 1$  tenemos que  $G(p) = I$  y considerando las condiciones iniciales de (3.10) encontramos la expresión de la curva geodésica:

$$X(t) = v_i t + p_i, \quad i = 1, \dots, n.$$

Esto es, las geodésicas son curvas  $x(t) : \mathbb{R} \rightarrow \mathbb{R}_{++}^n$ , definidas por:

$$x_t = (v_1 t + p_1, \dots, v_n t + p_n).$$

Observemos que la geodésica  $X(t)$  está definida para valores de  $t$  tal que  $v_i t + p_i > 0$ .

- Consideremos la variedad  $(\mathbb{R}_{++}^n, X^{-r})$ , donde las funciones  $h_i$  que definen la métrica están dadas por  $h_i(x_i) = x_i^{r/2}$ .

Resolvamos  $\int \frac{1}{x_i^{r/2}} dx_i = a_i t + b_i$ , para  $r \neq 2$ .

Por un lado tenemos que

$$\int \frac{1}{x_i^{r/2}} dx_i = x_i^{\frac{2-r}{2}},$$

así

$$x_i^{\frac{2-r}{2}} = a_i t + b_i,$$

lo que implica que

$$x_i = (a_i t + b_i)^{\frac{2}{2-r}}, \quad (3.11)$$

ahora bien,

$$x_i(0) = p_i \Leftrightarrow b_i^{\frac{2}{2-r}} = p_i \Leftrightarrow b_i = p_i^{\frac{2-r}{2}}.$$

Por otro lado,

$$x'_i(t) = \frac{2a_i}{2-r} (a_i t + b_i)^{\frac{r}{2-r}},$$

luego

$$x'_i(0) = v_i \Leftrightarrow \frac{2a_i}{2-r} b_i^{\frac{r}{2-r}} = v_i \Leftrightarrow a_i = \frac{v_i(2-r)}{2b_i^{\frac{r}{2-r}}} = \frac{v_i(2-r)}{2p_i^{\frac{r}{2}}},$$

luego sustituyendo las expresiones de  $a_i$  y  $b_i$  en (3.11), obtenemos la única geodésica tal que  $x_i(0) = p$  y  $x'_i(0) = v$

$$x_i(t) = \left( \frac{v_i(2-r)}{2p_i^{\frac{r}{2}}} t + p_i^{\frac{2-r}{2}} \right)^{\frac{2}{2-r}}, \quad i = 1, 2, \dots, n. \quad (3.12)$$

Si  $r = 2$ ,

$$x_i(t) = p_i e^{\left(\frac{v_i t}{p_i}\right)}, \quad i = 1, 2, \dots, n.$$

- Consideremos la variedad  $(\mathbb{R}_{++}^n, X^{-2}(I - X)^{-2})$ , donde las funciones  $h_i$  que definen la métrica están dadas por  $h_i(x_i) = x_i(1 - x_i)$ .

Tenemos que,

$$\int \frac{1}{x_i(1-x_i)} dx_i = \ln\left(\frac{x_i}{1-x_i}\right),$$

luego

$$\ln\left(\frac{x_i}{1-x_i}\right) = a_i t + b_i,$$

lo que implica que

$$x_i = \frac{e^{(a_i t + b_i)}}{1 + e^{(a_i t + b_i)}}, \quad (3.13)$$

ahora bien,

$$x_i(0) = p_i \Leftrightarrow \frac{e^{b_i}}{1 + e^{b_i}} = p_i \Leftrightarrow b_i = \ln\left(\frac{p_i}{1 - p_i}\right).$$

Por otro lado,

$$x'_i(t) = \frac{a_i e^{(a_i t + b_i)}}{(1 + e^{(a_i t + b_i)})^2},$$

luego

$$x'_i(0) = v_i \Leftrightarrow \frac{a_i e^{b_i}}{(1 + e^{b_i})^2} = v_i \Leftrightarrow a_i = \frac{v_i}{p_i(1 - p_i)},$$

luego sustituyendo las expresiones de  $a_i$  y  $b_i$  en (3.13), obtenemos que

$$x_i(t) = \frac{e^z}{1 + e^z}, \text{ donde } z = \frac{v_i}{p_i(1 - p_i)}t + \ln\left(\frac{p_i}{1 - p_i}\right) \quad i = 1, 2, \dots, n.$$

Por otro lado,

$$\frac{e^z}{1 + e^z} = \frac{1}{2} \left(1 + \frac{e^{z/2} - e^{-z/2}}{e^{z/2} + e^{-z/2}}\right) = \frac{1}{2}(1 + \tanh(z/2)),$$

ya que  $\tanh(w) = (e^w - e^{-w})/(e^w + e^{-w})$ , así la única geodésica tal que

$x_i(0) = p$  y  $x'_i(0) = v$  esta dada por :

$$x_i(t) = \frac{1}{2} \left(1 + \tanh\left(\frac{1}{2} \frac{v_i}{p_i(1 - p_i)}t + \frac{1}{2} \ln\left(\frac{p_i}{1 - p_i}\right)\right)\right) \quad i = 1, 2, \dots, n. \quad (3.14)$$

#### 4. Curvatura.

Recordemos que si en las coordenadas  $(U, \mathcal{X})$  escribimos  $X = \sum_{i=1}^n u^i X_i$ ,

$Y = \sum_{j=1}^n v^j X_j$ ,  $Z = \sum_{k=1}^n w^k X_k$ , obtenemos :

$$K(X, Y)Z = \sum_{i,j,k,l=1}^n K_{ijk}^l u^i v^j w^k X_l. \quad (3.15)$$

De la definición de curvatura tenemos:

$$k(X_i, X_j)X_k = \nabla_{X_j}(\nabla_{X_i}X_k) - \nabla_{X_i}(\nabla_{X_j}X_k) + \nabla_{[X_i, X_j]}X_k,$$

como la conexión es de Levi Civita se tiene  $[X_i, X_j] = 0$ . Así,

$$k(X_i, X_j)X_k = \nabla_{X_j}(\nabla_{X_i}X_k) - \nabla_{X_i}(\nabla_{X_j}X_k).$$

Si  $i = j$ , entonces  $K(X_i, X_j)X_k = 0$ . Supongamos que  $i \neq j$ . Sabemos que:

$$\nabla_{X_i}X_k = \sum_{j=1}^n \Gamma_{ik}^j X_j.$$

Sustituyendo los símbolos de Christoffel tenemos:

$$\nabla_{X_i}X_k = \sum_{j=1}^n \left( -\frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \delta_{ij} \delta_{ik} \right) X_j = -\frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \delta_{ik} X_i, \quad (3.16)$$

luego aplicando  $\nabla_{X_j}$  se tiene:

$$\nabla_{X_j}(\nabla_{X_i}X_k) = \nabla_{X_j} \left( -\frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \delta_{ik} X_i \right),$$

usando la propiedad iii) de la conexión resulta:

$$\nabla_{X_j}(\nabla_{X_i}X_k) = -\frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \delta_{ik} \nabla_{X_j}X_i + X_j \left( -\frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \delta_{ik} \right) X_i.$$

Usando (3.16) y teniendo en cuenta que  $i \neq j$ , los dos términos de la suma anterior son iguales a cero y así

$$\nabla_{X_j}(\nabla_{X_i}X_k) = 0.$$

Haciendo un razonamiento similar al anterior se obtiene  $\nabla_{X_i}(\nabla_{X_j}X_k) = 0$

y por tanto

$$k(X_i, X_j)X_k = 0, \quad \forall i, j, k = 1, 2, \dots, n,$$

y así

$$k(X, Y)Z = 0. \quad (3.17)$$

Concluimos que las variedades riemannianas  $\mathbb{R}_{++}^n$  y  $\mathcal{C}_0^n$  con métrica diagonal tienen curvatura cero.

### 5. Gradiente y Hessiana.

Teniendo en cuenta la observación (1.7.3) el gradiente de la variedad riemanniana  $\mathbb{R}_{++}^n$  con métrica  $G(x) = \text{diag}\left(\frac{1}{h_1(x_1)^2}, \frac{1}{h_2(x_2)^2}, \dots, \frac{1}{h_n(x_n)^2}\right)$ , viene dado por:

$$\text{grad}f(x) = G^{-1}(x)f'(x) = \text{diag}((h_1(x_1))^2, (h_2(x_2))^2, \dots, (h_n(x_n))^2)f'(x).$$

#### Ejemplo 3.0.4

- Si  $h_i(x_i) = 1$ , entonces  $G(x) = I$  y así el gradiente riemanniano coincide con el gradiente usual.
- Si  $h_i(x_i) = x_i$ , entonces  $\text{grad}f(x) = X^2 f'(x)$
- Si  $h_i(x_i) = x_i^{r/2}$ , entonces  $\text{grad}f(x) = X^r f'(x)$
- Si  $h_i(x_i) = x_i(1 - x_i)$ , entonces  $\text{grad}f(x) = X^2(I - X_i)^2 f'(x)$

Por otro lado, de la observación (1.7.4) tenemos que:

$$\langle H_p^f X_i, X_j \rangle = \left( \frac{\partial^2 f}{\partial x_i \partial x_j} - \sum_{m=1}^n \Gamma_{ij}^m X_m \right) f,$$

sustituyendo los símbolos de Christoffel obtenemos:

$$\begin{aligned} \langle H_p^f X_i, X_j \rangle &= \left( \frac{\partial^2 f}{\partial x_i \partial x_j} + \sum_{m=1}^n \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \delta_{im} \delta_{ij} X_m \right) f \\ &= \left( \frac{\partial^2 f}{\partial x_i \partial x_j} + \frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \delta_{ij} X_i \frac{\partial}{\partial x_i} \right) f. \end{aligned}$$

Así, esta es la matriz que representa la Hessiana de la función  $f$ , la cual se puede representar como:

$$H_x^f = f''(x) + G(x)^{1/2} (G(x)^{-1/2})' F(x)',$$

donde:

$$F'(x) = \text{diag} \left( \frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n} \right),$$

$$G(x) = \text{diag} \left( \frac{1}{h_1(x_1)^2}, \frac{1}{h_2(x_2)^2}, \dots, \frac{1}{h_n(x_n)^2} \right),$$

$$f''(x) = \text{diag} \left( \frac{\partial^2 f}{\partial x_1^2}, \frac{\partial^2 f}{\partial x_2^2}, \dots, \frac{\partial^2 f}{\partial x_n^2} \right)$$

### Ejemplo 3.0.5

- Si  $h_i(x_i) = 1$ , entonces  $G(x) = I$  y así el hessiana riemanniana coincide con el hessiana usual.
- Si  $h_i(x_i) = x_i$ , entonces  $H_x^f = f''(x) + X^{-1} F'(x)$
- Si  $h_i(x_i) = x_i^{r/2}$ , entonces  $H_x^f = f''(x) + \frac{r}{2} X^{-1} F'(x)$

**Observación 3.0.1** *De manera similar, el gradiente y la hessiana de  $f$  en la variedad  $C_0^n$  tienen las mismas expresiones dadas en arriba.*

6. Distancia Riemanniana.

De la definición 1.7.6, tenemos que la distancia riemanniana se obtiene mediante

$$d(p, q) = \inf_{\gamma} \int_a^b \|\gamma'(t)\| dt,$$

con  $\gamma(0) = p$  y  $\gamma(t_0) = q$ . Veamos algunos ejemplos:

**Ejemplo 3.0.6** ■

*Consideremos la variedad  $(\mathbb{R}_{++}^n, X^{-2})$ . La única geodésica tal que  $x_i(0) = p_i$  y  $x_i(t_0) = q_i$ ,  $t_0 > 0$ , viene dada por :*

$$x_i(t) = p_i \left(\frac{q_i}{p_i}\right)^{t/t_0}, \quad i = 1, 2, \dots, n.$$

*Por otro lado,*

$$x'_i(t) = \frac{p_i}{t_0} \ln\left(\frac{q_i}{p_i}\right) \left(\frac{q_i}{p_i}\right)^{t/t_0}.$$

*Ahora bien,*

$$\frac{x'_i(t)}{x_i(t)} = \frac{1}{t_0} \ln\left(\frac{q_i}{p_i}\right),$$

*lo que implica que,*

$$\left(\frac{x'_i(t)}{x_i(t)}\right)^2 = \frac{1}{t_0^2} \ln^2\left(\frac{q_i}{p_i}\right),$$



luego,

$$\begin{aligned}
 d(p, q) &= \int_0^{t_0} \|x'(t)\| dt \\
 &= \int_0^{t_0} \langle x'(t), x'(t) \rangle_{x(t)}^{1/2} dt \\
 &= \int_0^{t_0} \left( \sum_{i=1}^n \frac{x'_i(t)x'_i(t)}{x_i^2(t)} \right)^{1/2} dt \\
 &= \int_0^{t_0} \left( \sum_{i=1}^n \frac{x_i'^2(t)}{x_i^2(t)} \right)^{1/2} dt \\
 &= \int_0^{t_0} \left( \sum_{i=1}^n \frac{1}{t_0^2} \ln^2 \left( \frac{q_i}{p_i} \right) \right)^{1/2} dt \\
 &= \left( \sum_{i=1}^n \frac{1}{t_0^2} \ln^2 \left( \frac{q_i}{p_i} \right) \right)^{1/2} \int_0^{t_0} dt \\
 &= \left( \sum_{i=1}^n \ln^2 \left( \frac{q_i}{p_i} \right) \right)^{1/2}
 \end{aligned}$$

$$\text{Así, } d(p, q) = \left( \sum_{i=1}^n \ln^2 \left( \frac{q_i}{p_i} \right) \right)^{1/2}$$

Consideremos la variedad  $(\mathbb{R}_{++}^n, X^{-2}(I - X)^{-2})$ . La única geodésica tal que

$x_i(0) = p_i$  y  $x_i(t_0) = q_i$ ,  $t_0 > 0$ , está dada por :

$$x_i(t) = \frac{1}{2} \left\{ 1 + \tanh \left( \frac{1}{2} \left[ \ln \left( \frac{q_i}{1 - q_i} \right) - \ln \left( \frac{p_i}{1 - p_i} \right) \right] \frac{t}{t_0} + \ln \left( \frac{p_i}{1 - p_i} \right) \right) \right\} \quad i = 1, \dots, n.$$

Por un lado,

$$x'_i(t) = \frac{1}{4} \frac{1}{t_0} \left[ \ln \left( \frac{q_i}{1 - q_i} \right) - \ln \left( \frac{p_i}{1 - p_i} \right) \right] \operatorname{sech}^2 \left( \frac{1}{2} \left[ \ln \left( \frac{q_i}{1 - q_i} \right) - \ln \left( \frac{p_i}{1 - p_i} \right) \right] \frac{t}{t_0} + \ln \left( \frac{p_i}{1 - p_i} \right) \right),$$

luego,

$$\frac{x'_i(t)}{x_i(t)} = \frac{1}{t_0} \left[ \ln \left( \frac{q_i}{1 - q_i} \right) - \ln \left( \frac{p_i}{1 - p_i} \right) \right],$$

así,

$$\begin{aligned}
 d(p, q) &= \int_0^{t_0} \|x'(t)\| dt \\
 &= \int_0^{t_0} \left\{ \sum_{i=1}^n \frac{1}{t_0^2} \left[ \ln\left(\frac{q_i}{1-q_i}\right) - \ln\left(\frac{p_i}{1-p_i}\right) \right]^2 \right\}^{1/2} dt \\
 &= \left\{ \sum_{i=1}^n \frac{1}{t_0^2} \left[ \ln\left(\frac{q_i}{1-q_i}\right) - \ln\left(\frac{p_i}{1-p_i}\right) \right]^2 \right\}^{1/2} \int_0^{t_0} dt \\
 &= \left\{ \sum_{i=1}^n \left[ \ln\left(\frac{q_i}{1-q_i}\right) - \ln\left(\frac{p_i}{1-p_i}\right) \right]^2 \right\}^{1/2}.
 \end{aligned}$$

$$\text{Así, } d(p, q) = \left\{ \sum_{i=1}^n \left[ \ln\left(\frac{q_i}{1-q_i}\right) - \ln\left(\frac{p_i}{1-p_i}\right) \right]^2 \right\}^{1/2}.$$

---

## Capítulo 4

### Experimentos Numéricos

---

En este capítulo se presentan experimentos numéricos para resolver problemas de optimización en variedades riemannianas. De manera específica, en primer lugar vamos a presentar el algoritmo del gradiente con regla de Armijo generalizada, desarrollado por Papa Quiroz en [9] y en segundo lugar, vamos a proponer un algoritmo basado en la búsqueda no monótona de  $\mathbb{R}^n$ .

El problema que es de nuestro interés resolver, es el siguiente:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a.} \quad & 0 \leq x_i \leq 1, \quad i = 1, 2, \dots, n. \end{aligned}$$

donde  $f$  es una función cuasiconvexa.

Tomando la variedad riemanniana  $((0, 1)^n, X^{-2}(I - X)^{-2})$  se tiene que:

- a)** Es completa (geodésicamente), pues dado un punto cualquiera  $x \in M$  y una dirección cualquiera  $v \in T_x M$ , por (3.14) tenemos que la única geodésica tal

que  $x_i(0) = p$  y  $x'_i(0) = v$  esta dada por :

$$x_i(t) = \frac{1}{2} \left\{ 1 + \tanh \left( \frac{1}{2} \frac{v_i}{p_i(1-p_i)} t + \frac{1}{2} \ln \left( \frac{p_i}{1-p_i} \right) \right) \right\}, \quad i = 1, 2, \dots, n,$$

la cual está definida para todo  $t$ .

- b) Es convexa , ya que dados  $p, q \in (0, 1)^n$  existe una única geodésica  $x : [0, 1] \rightarrow (0, 1)^n$  uniendo a  $p$  y  $q$ , donde  $x(0) = p$  y  $x(1) = q$ , la cual está definida para todo  $t$  y está dada por:

$$x_i(t) = \frac{1}{2} \left\{ 1 + \tanh \left( \frac{1}{2} \left[ \ln \left( \frac{q_i}{1-q_i} \right) - \ln \left( \frac{p_i}{1-p_i} \right) \right] t + \ln \left( \frac{p_i}{1-p_i} \right) \right) \right\} \quad i = 1, 2, \dots, n.$$

En efecto, de (3.13) tenemos:

$$x_i = \frac{e^{(a_i t + b_i)}}{1 + e^{(a_i t + b_i)}}, \quad (4.1)$$

haciendo  $x_i(0) = p_i$  y  $x_i(1) = q_i$  se obtiene que:

$$b_i = \ln \left( \frac{p_i}{1-p_i} \right) \text{ y } a_i = \ln \left( \frac{q_i}{1-q_i} \right) - \ln \left( \frac{p_i}{1-p_i} \right), \text{ sustituyendo esto en (4.1) y teniendo}$$

en cuenta que  $\tanh(w) = (e^w - e^{-w}) / (e^w + e^{-w})$ , se obtiene el resultado.

- c) Tiene curvatura nula (ver 3.17).

## 4.1. Búsqueda de Armijo Generalizada

Se presenta el algoritmo del gradiente con búsqueda de Armijo generalizada desarrollado en [9], para el caso particular de métricas diagonales.

**Algoritmo 4.1.1**

1. Dada una tolerancia  $\varepsilon > 0$  y  $x^k = (x_1^k, x_2^k, \dots, x_n^k)$ , calcule la métrica

$$G(x^k) = X_k^{-2}(I - X_k)^{-2}, \text{ donde } X_k = \text{diag}(x_1^k, x_2^k, \dots, x_n^k).$$

2. Calcular la dirección de descenso:

$$d^k = -G(x^k)^{-1} f'(x^k).$$

3. La única geodésica  $x(t)$ ,  $t \geq 0$ , tal que  $x(0) = p$  y  $x'(0) = d^k$  esta dada por :

$$x_i(t) = \frac{1}{2} \left\{ 1 + \tanh \left( \frac{1}{2} \frac{d_i^k}{p_i(1-p_i)} t + \frac{1}{2} \ln \left( \frac{p_i}{1-p_i} \right) \right) \right\}, \quad i = 1, 2, \dots, n,$$

4. Elegir  $t_k = 2^{-ik}$ , donde  $ik$  es el menor número natural tal que:

$$f(x(t_k)) \leq f(x^k) - \beta t_k \|d^k\|^2, \quad (4.2)$$

donde  $\beta \in (0, 1)$ .

5. Hacer  $x^{k+1} = x^k(t_k)$ , y calcular la distancia geodésica entre los puntos  $x^k$  y  $x^{k+1}$ , como:

$$d(x^k, x^{k+1}) = \left\{ \sum_{i=1}^n \left[ \ln \left( \frac{x_i^{k+1}}{1-x_i^{k+1}} \right) - \ln \left( \frac{x_i^k}{1-x_i^k} \right) \right]^2 \right\}^{1/2}$$

6. Criterio de parada: si  $|d(x^k, x^{k+1})| < \varepsilon$  parar, caso contrario hacer  $x^k \leftarrow x^{k+1}$  y volver al paso 1.

## 4.2. Búsqueda no monótona

Se presenta el siguiente algoritmo para resolver problemas de optimización en variedades de Riemann, basado en la búsqueda lineal no monótona de  $\mathbb{R}^n$  descrita en capítulo 1. Recordemos el tamaño de paso  $\alpha_k$  es determinado de la siguiente manera:

Dado  $a > 0$ ,  $\theta \in (0, 1)$  y  $\beta \in (0, 1)$ ,  $\alpha_k = \theta^{h_k} a$ , donde  $h_k$  es el primer entero no negativo  $h$  para el cual:

$$f(x^k + \alpha_k d^k) \leq \max_{0 \leq j \leq m(k)} [f(x^{k-j})] + \beta \theta^h a \nabla f(x^k)^T d^k,$$

donde  $m(0) = 0$  y  $0 \leq m_k \leq \min\{m(k-1) + 1, M\}$ ,  $k \geq 1$ .

Para el caso de variedades riemannianas el algoritmo es el siguiente:

### Algoritmo 4.2.1

#### *Método de máximo descenso con búsqueda no monótona*

1. Dado  $x^k$ ,  $k \geq 1$ , un punto factible, calcular la dirección  $d^k = -\text{grad}f(x^k)$  (gradiente riemanniano).
2. Determinar la única geodésica  $x(t)$  con  $t \geq 0$  de  $M$  tal que  $x(0) = x^k$  y  $x'(0) = d^k$ .

3. Hacer,

$$x^{k+1} = x(t_k),$$

donde el tamaño del paso  $t_k$  es determinado de la siguiente manera:

Elegir  $t_k = 2^{-i_k}$ , donde  $i_k$  es el menor número natural tal que

$$f(x(t)) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x^{k-j}) - \alpha t_k \|\text{grad}f(x^k)\|^2, \quad \alpha \in (0, 1),$$

Usando la variedad riemanniana  $((0, 1)^n, X^{-2}(I - X)^{-2})$  y para el caso particular de métricas diagonales, se presenta el siguiente algoritmo:

#### Algoritmo 4.2.2

1. Dada una tolerancia  $\varepsilon > 0$  y  $x^k = (x_1^k, x_2^k, \dots, x_n^k)$ , calcule la métrica

$$G(x^k) = X_k^{-2}(I - X_k)^{-2}, \quad \text{donde } X_k = \text{diag}(x_1^k, x_2^k, \dots, x_n^k).$$

2. Calcular la dirección de descenso:

$$d^k = -G(x^k)^{-1} f'(x^k).$$

3. La única geodésica  $x(t)$ ,  $t \geq 0$ , tal que  $x(0) = p$  y  $x'(0) = d^k$  esta dada por :

$$x_i(t) = \frac{1}{2} \left\{ 1 + \tanh \left( \frac{1}{2} \frac{d_i^k}{p_i(1-p_i)} t + \frac{1}{2} \ln \left( \frac{p_i}{1-p_i} \right) \right) \right\}, \quad i = 1, 2, \dots, n,$$

4. Elegir  $t_k = 2^{-i_k}$ , donde  $i_k$  es el menor número natural tal que:

$$f(x(t_k)) \leq \max_{0 \leq j \leq \min\{k, M-1\}} f(x^{k-j}) - \beta t_k^2 \|d^k\|^2, \quad (4.3)$$

donde  $\beta \in (0, 1)$ .

5. Hacer  $x^{k+1} = x^k(t_k)$ , y calcular la distancia geodésica entre los puntos  $x^k$  y  $x^{k+1}$ , como:

$$d(x^k, x^{k+1}) = \left\{ \sum_{i=1}^n \left[ \ln \left( \frac{x_i^{k+1}}{1 - x_i^{k+1}} \right) - \ln \left( \frac{x_i^k}{1 - x_i^k} \right) \right]^2 \right\}^{1/2}$$

6. Criterio de parada: si  $|d(x^k, x^{k+1})| < \varepsilon$  parar, caso contrario hacer  $x^k \leftarrow x^{k+1}$  y volver al paso 1.



A continuación se presentan experimentos con tres funciones cuasiconvexas utilizando los algoritmos (4.1.1) y (4.2.2), el código fue implementado en matlab. Para los experimentos 1 y 2 fueron utilizados los siguientes parámetros:  $\varepsilon = 0,000001$ ,  $\beta = 0,9$ ,  $M = 5$ ,  $t_0 = 3$ , y para el experimento 3,  $\varepsilon = 0,000001$ ,  $\beta = 0,95$ ,  $M = 5$ ,  $t_0 = 2$ . En la tablas que se presentan,  $X^0$  denota el punto inicial del algoritmo, Iter. denota el número de iteraciones, L.A. denota el número de veces que se utiliza Armijo, BNM las veces que se utiliza la búsqueda no monótona, P.Óptimo denota la aproximación del punto óptimo, V.Óptimo denota la aproximación del valor óptimo y finalmente, D.Riemann denota la distancia riemanniana entre dos iteraciones consecutivas.

**Experimento 1**

La primera función a utilizar es  $f(x, y) = \sqrt{-\ln[(x - x^2)(y - y^2)]}$ , la cual es cuasiconvexa en  $((0, 1)^n, X^{-2}(I - X)^{-2})$  y tiene un único mínimo en  $x^*=(0.5,0.5)$  con valor óptimo  $f^* = 2\sqrt{\ln(2)} = 1,665109222$ .

**Tabla 1 (Búsqueda de Armijo Generalizada)**

$x^0$	Iter	L.A.	P.Óptimo	V.Óptimo	D.Riemann
(0.45,0.51)	71	71	(0,499999, 0,500000)	1.665109	8.904980e-007
(0.4,0.6)	77	77	(0,499999, 0,500000)	1.665109	9.542784e-007
(0.1,0.9)	86	84	(0,499999, 0,500000)	1.665109	9.728218e-007
(0.2,0.3)	84	84	(0,499999, 0,499999)	1.665109	9.944773e-007
(0.7,0.6)	80	80	(0,500000, 0,500000)	1.665109	9.980098e-007

**Tabla 2 (Búsqueda no monótona)**

$x^0$	Iter	BNM	P.Óptimo	V.Óptimo	D.Riemann
(0.45,0.51)	23	4	(0,499999, 0,500000)	1.665109	5.549691e-007
(0.4,0.6)	24	4	(0,499999, 0,500000)	1.665109	8.705161e-007
(0.1,0.9)	30	7	(0,499999, 0,500000)	1.665109	8.416977e-007
(0.2,0.3)	26	4	(0,499999, 0,499999)	1.665109	8.888718e-007
(0.7,0.6)	25	4	(0,500000, 0,500000)	1.665109	8.232247e-007

**Experimento 2**

La segunda función que utilizaremos  $f(x, y) = \ln[1 - \ln[(x-x^2)(y-y^2)]]$ , la cual es cuasiconvexa en  $((0, 1)^n, X^{-2}(I - X)^{-2})$  y tiene un único mínimo en  $x^*=(0.5,0.5)$  con valor óptimo  $f^* = \ln(1 + 4 \ln(2)) = 1,32776143$ .

**Tabla 3 (Búsqueda de Armijo Generalizada)**

$x^0$	Iter	L.A.	P.Óptimo	V.Óptimo	D.Riemann
(0.45,0.51)	80	80	(0,499999, 0,500000)	1.327761	9.379211e-007
(0.4,0.6)	87	87	(0,499999, 0,500000)	1.327761	9.893387e-007
(0.1,0.9)	97	94	(0,499999, 0,500000)	1.327761	9.926292e-007
(0.2,0.3)	94	93	(0,499999, 0,499999)	1.327761	8.942494e-007
(0.7,0.6)	91	91	(0,500000, 0,500000)	1.327761	9.605749e-007

**Tabla 4 (Búsqueda no monótona)**

$x^0$	Iter	BNM	P.Óptimo	V.Óptimo	D.Riemann
(0.45,0.51)	26	4	(0,499999, 0,500000)	1.327761	6.649036e-007
(0.4,0.6)	28	4	(0,499999, 0,500000)	1.327761	6.913453e-007
(0.1,0.9)	37	11	(0,499999, 0,500000)	1.327761	9.755524e-007
(0.2,0.3)	30	4	(0,499999, 0,499999)	1.327761	8.770180e-007
(0.7,0.6)	29	4	(0,500000, 0,500000)	1.327761	7.218485e-007

**Experimento 3**

Por ultimo, la función  $f(x, y) = \arctan[-\ln[(x - x^2)(y - y^2)]]$ , al igual que las anteriores es cuasiconvexa en  $((0, 1)^n, X^{-2}(I - X)^{-2})$  y tiene un único mínimo en  $x^*=(0.5,0.5)$  con valor óptimo  $f^* = \arctg(4 \ln(2)) = 1,224644415$ .

**Tabla 5 (Búsqueda de Armijo Generalizada)**

$x^0$	Iter	L.A.	P.Óptimo	V.Óptimo	D.Riemann
(0.45,0.51)	170	170	(0,499997, 0,500000)	1.224644	9.951586e-007
(0.4,0.6)	186	184	(0,499998, 0,500001)	1.224644	9.834920e-007
(0.1,0.9)	211	185	(0,499998, 0,500001)	1.224644	9.960921e-007
(0.2,0.3)	197	184	(0,499998, 0,499998)	1.224644	9.858996e-007
(0.7,0.6)	191	184	(0,500001, 0,500000)	1.224644	9.438545e-007

**Tabla 6 (Búsqueda no monótona)**

$x^0$	Iter	BNM	P.Óptimo	V.Óptimo	D.Riemann
(0.45,0.51)	84	2	(0,499998, 0,500000)	1.224644	9.289290e-007
(0.4,0.6)	93	2	(0,499998, 0,500001)	1.224644	8.953253e-007
(0.1,0.9)	117	2	(0,499998, 0,500001)	1.224644	9.587617e-007
(0.2,0.3)	104	2	(0,499998, 0,499999)	1.224644	8.952175e-007
(0.7,0.6)	97	2	(0,500001, 0,500000)	1.224644	9.700283e-007

## Conclusión

---

Existe una gran gama de métodos que se han desarrollado hasta la fecha, que resuelven problemas de optimización en variedades riemannianas; en este trabajo estudiamos las ideas principales de algunos de ellos, entre los cuales tenemos, el método de máximo descenso, con búsqueda geodésica exacta y con búsqueda de Armijo, una regularización proximal del método de máximo descenso, entre otros. Usando las métricas diagonales en el hipercubo , pudimos mostrar ejemplos numéricos para funciones específicas, donde la búsqueda no monótona es más eficiente que la búsqueda de Armijo. Se harán estudios de convergencia para esta busca no monótona, así como también se realizará una mayor experimentación con otras búsquedas; además se pretende estudiar la posibilidad de usar el método de barzilai-borwein.

## Apéndice

---

### Código del Algoritmo: Búsqueda de Armijo Generalizada.

```
syms x y
xk = input('Introduzca el punto inicial  $x^k$  ([ ; ])\');
fun = input('Introduzca la función f (Variables x,y)\');
grad = input('Introduzca el gradiente de la función f ([ ; ])\');
n = 1;
itermax = 1000;
cont2 = 0;
while n <= itermax;
dxk = diag(xk);
Gk = dxk^(-2) * ((eye(size(dxk)) - dxk))^(-2);
gradant = subs(grad, x, y, xk(1), xk(2));
dk = -dxk^(2) * ((eye(size(dxk)) - dxk))^(2) * gradant;
a = dk(1);
b = dk(2);
```

```

c = xk(1);
d = xk(2);
e = gradant(1);
f = gradant(2);
t = 3;
xt1 = 1/2 * (1 + tanh(-1/2 * c * (1 - c) * e * t + 1/2 * log(c/(1 - c))));
xt2 = 1/2 * (1 + tanh(-1/2 * d * (1 - d) * f * t + 1/2 * log(d/(1 - d))));
dist = sqrt((log(xt1/(1 - xt1)) - log(c/(1 - c)))^2 + (log(xt2/(1 - xt2)) - log(d/(1 -
d)))^2);
if dist < 0,000001
    disp('Se cumple la tolerancia')
    break
else
    fxk = subs(fun, x, y, xt1, xt2);
    fxk = subs(fun, x, y, c, d);
    normadk = (a^2/(c^2 * (1 - c)^2)) + (b^2/(d^2 * (1 - d)^2));
    cont1 = 0;
    while fxk > fxk - (0,9) * t * normadk
        t = 2^(-cont1);
        xt1 = 1/2 * (1 + tanh(-1/2 * c * (1 - c) * e * t + 1/2 * log(c/(1 - c))));

```

```
xt2 = 1/2 * (1 + tanh(-1/2 * d * (1 - d) * f * t + 1/2 * log(d/(1 - d))));  
fxt = subs(fun, x, y, xt1, xt2);  
cont1 = cont1 + 1;  
end  
cont2 = cont2 + cont1;  
xk(1) = xt1;  
xk(2) = xt2;  
n = n + 1;  
end  
end  
disp('El mínimo es alcanzado en ')  
disp(xk)  
disp('El valor de f es')  
f = subs(fun, x, y, xk(1), xk(2));  
disp(f)  
disp('N de iteraciones realizadas')  
iter = n - 1;  
disp(iter)  
disp('Llamadas de armijo')  
disp(cont2)
```



```
disp('Distancia riemanniana')
```

```
disp(dist)
```

**Código del Algoritmo: Búsqueda no Monótona.**

```
syms x y
```

```
xk = input('Introduzca el punto inicial  $x^k$  ([ ; ])' );
```

```
fun = input('Introduzca la función f (Variables x,y)' );
```

```
grad = input('Introduzca el gradiente de la función f ([ ; ])' );
```

```
m = [0 0 0 0 0];
```

```
n = 1;
```

```
i = 1;
```

```
itermax = 1000;
```

```
cont2 = 0;
```

```
while n <= itermax;
```

```
dxk = diag(xk);
```

```
Gk = dxk^(-2) * ((eye(size(dxk)) - dxk))^(-2);
```

```
gradant = subs(grad, x, y, xk(1), xk(2));
```

```
dk = -dxk^2 * ((eye(size(dxk)) - dxk))^2 * gradant;
```

```
a = dk(1);
```

```
b = dk(2);
```

```
c = xk(1);  
d = xk(2);  
e = gradant(1);  
f = gradant(2);  
t = 3;  
xt1 = 1/2 * (1 + tanh(-1/2 * c * (1 - c) * e * t + 1/2 * log(c/(1 - c))));  
xt2 = 1/2 * (1 + tanh(-1/2 * d * (1 - d) * f * t + 1/2 * log(d/(1 - d))));  
dist = sqrt((log(xt1/(1 - xt1)) - log(c/(1 - c)))^2 + (log(xt2/(1 - xt2)) - log(d/(1 -  
d)))^2);  
if dist < 0,000001  
disp('Se cumple la tolerancia')  
break  
else  
fmt = subs(fun, x, y, xt1, xt2);  
fxk = subs(fun, x, y, c, d);  
normadk = (a^2/(c^2 * (1 - c)^2)) + (b^2/(d^2 * (1 - d)^2));  
if i <= 5;  
m(i) = fxk;  
else  
i = i - 5;
```

```
m(i) = f xk;
end
cont1 = 0;
while fxt > max(m) - (0,9) * t^2 * normadk
t = 2^(-cont1);
xt1 = 1/2 * (1 + tanh(-1/2 * c * (1 - c) * e * t + 1/2 * log(c/(1 - c))));
xt2 = 1/2 * (1 + tanh(-1/2 * d * (1 - d) * f * t + 1/2 * log(d/(1 - d))));
fxt = subs(fun, x, y, xt1, xt2);
cont1 = cont1 + 1;
end
cont2 = cont2 + cont1;
xk(1) = xt1;
xk(2) = xt2;
n = n + 1;
i = i + 1;
end
end
disp('El mínimo es alcanzado en ')
disp(xk)
disp('El valor de f es')
```

```
f = subs(fun, x, y, xk(1), xk(2));
```

```
disp(f)
```

```
disp('N de iteraciones realizadas')
```

```
iter = n - 1;
```

```
disp(iter)
```

```
disp('Veces que se realiza la búsqueda no monótona')
```

```
disp(cont2)
```

```
disp('Distancia riemanniana')
```

```
disp(dist)
```

Función 1:

```
sqrt(-log((x - x^2) * (y - y^2)))
```

Gradiente:

```
[((x - 1/2)/(x - x^2)) * (1/(sqrt(-log((x - x^2) * (y - y^2)))))); ((y - 1/2)/(y - y^2)) * (1/(sqrt(-log((x - x^2) * (y - y^2)))))]
```

Función 2:

```
log(1 - log((x - x^2) * (y - y^2)))
```

Gradiente:

```
[((2 * x - 1)/(x - x^2)) * (1/((1 - log((x - x^2) * (y - y^2))))); ((2 * y - 1)/(y - y^2)) * (1/((1 - log((x - x^2) * (y - y^2)))))]
```

## Bibliografía

---

- [1] D.LUENBERGER, *The Gradient Projection Method Along Geodesic*, Management Science. Vol 18, N° 11 (1972).
- [2] D.GABAY, *Minimizing a Differentiable Function over a Differential Manifold*, Rapports de Recherche, N° 9 (1980).
- [3] J.NETO, L. DE LIMA AND P. OLIVEIRA, *Geodesic Algorithms in Riemannian Geometry*, Balkan Journal of Geometry and its Applications, Vol 3, N° 2, (1998) pp.89-100.
- [4] J.NETO, O. FERREIRA AND L. PEREZ, *A proximal Regularization of the steepest Descent Method in Riemannian Manifold*, Balkan Journal of Geometry and its Applications, Vol 4, N° 2 (1999) pp.1-8.
- [5] E. QUIROZ, E. QUISPE AND P. OLIVEIRA, *steepest Descent Method for Quasiconvex Minimization on Riemannian Manifold*, Federal University of Rio de Janeiro, PESC-COPPE (2006).

- [6] S. SOUSA, P. SANTOS AND J. NETO, *A Descent Algorithm for Quasi-convex Optimization Over Riemannian Manifold*, XLIISBPO, Universidade Federal do Piauí (2010).
- [7] S. DO CARMO, MANFREDO PERDIGAO, *Geometria Diferencial de Curvas e Superfícies*, Rio de Janeiro: Sociedade Brasileira de Matemática, terceira edición (2005).
- [8] S. DO CARMO, MANFREDO PERDIGAO, *Geometria Riemanniana*, Rio de Janeiro: IMPA-Projeto Euclides, segunda edición, (1988).
- [9] ERIK QUIROZ, *Un Método no Euclidiano para Problemas de Optimización en Espacios Euclidianos*, Universidad Nacional Del Callao (2007).
- [10] E. QUIROZ AND P. OLIVEIRA, *New Results On Linear Optimization Through Diagonal Metrics And Riemannian Geometry Tools*, Federal University of Rio de Janeiro, PESC-COPPE (2004).
- [11] L. GRIPO, F. LAMPARIELLO AND S. LUCIDI, *A Nonmonotone Line Search Technique For Newton's Method*, Istituto di analisi dei Sistemi ed Informatica del CNR (1985).
- [12] F. CAMARGO, *Estudo Comparativo de Passos Espectrais e Buscas Lineares não monótonas*, Instituto De matemática E Estatística Da Universidade De São Paulo (2008).

- [13] J. NOCEDAL AND S. WRIGHT, *Numerical Optimization* , Springer Series in Operations Research.