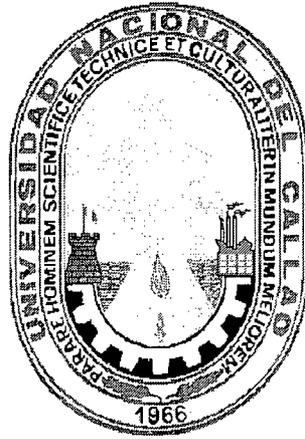


t
510
09

UNIVERSIDAD NACIONAL DEL CALLAO

FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

ESCUELA PROFESIONAL DE MATEMÁTICA



**EL MÉTODO DE MÁXIMO DESCENSO PARA
FUNCIONES CUASI-CONVEXAS EN VARIEDADES
RIEMANNIANAS**

**TESIS PARA OPTAR EL TÍTULO PROFESIONAL DE
LICENCIADO EN MATEMÁTICA**

ELSA MARISA QUISPE CÁRDENAS

CALLAO B - PERÚ

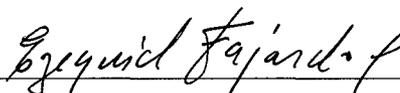
ENERO - 2008

EL MÉTODO DE MÁXIMO DESCENSO PARA FUNCIONES
CUASI-CONVEXAS EN VARIEDADES RIEMANNIANAS

Elsa Marisa Quispe Cárdenas

Tesis presentada a consideración del Cuerpo Docente, de la Facultad de Ciencias Naturales y Matemática de la Universidad Nacional del Callao, como parte de los requisitos para obtener el Título Profesional de Licenciado en Matemática.

Aprobada por :



Prof. Ezequiel Fajardo Campos, Lic.

Presidente



Prof. Moisés Simón Lázaro Carrión, Lic.

Secretario



Prof. Wilfredo Mendoza Quispe, Lic.

Vocal



Prof. José Raúl Luyo Sánchez, Dr.

Suplente



Prof. Erik Alex Papa Quiroz, DSc.

Asesor

Callao - Perú

Enero - 2008

FICHA CATALOGRÁFICA

QUISPE CÁRDENAS, ELSA MARISA

El método de máximo descenso para funciones cuasi-convexas en variedades riemannianas, Callao [2008].

ix, 71p., 29.7 cm. (UNAC, Licenciado en Matemática, 2008).

Tesis, Universidad Nacional del Callao, Facultad de Ciencias Naturales y Matemática. I. Matemática.

I. UNAC/FCNM II. Título (Serie).

A: juan-pa.

Agradecimientos

A todas las personas que cercanas a mi, hicieron de los momentos en que desarrollé ésta tesis: calmados, divertidos y estresantes en los últimos momentos, a unos por darme ánimos lanzandome palabras alentadoras, dado que, ya me había cansado de encontrar los α_i y confundirme con los x_j , a otros, por decir francamente!! sandeses, haciendome reir o reaccionar con corage por momentos, y a otros más, por dejar fluir ideas y sensaciones prohibidas, para plasmar esta tesis, finalmente, también a los que ni les va y ni les viene. Muchas gracias a todos!!.

RESUMEN

EL MÉTODO DE MÁXIMO DESCENSO PARA FUNCIONES CUASI-CONVEXAS EN VARIEDADES RIEMANNIANAS

Elsa Marisa Quispe Cárdenas

Enero - 2008

Asesor: DSc. Erik Alex Papa Quiroz

Título Obtenido: Licenciado en Matemática

En este trabajo, probamos la convergencia global del método del máximo descenso con búsqueda generalizada de Armijo para resolver problemas de minimización con funciones objetivo cuasi-convexas definidas en una variedad riemanniana completa con curvatura seccional no negativa. Resultados de convergencia obtenidos en espacios euclidianos, llegan a ser casos particulares de este desarrollo. Además, introducimos una clase de métricas diagonales en la variedad \mathbb{R}_{++}^n y estudiamos sus propiedades geométricas, como son: geodésicas, curvatura, distancias riemannianas, etc.

Palabras Claves:

Variedad riemanniana.

Método de máximo descenso.

Funciones cuasi-convexa.

Convergencia global.

ABSTRACT

**STEEPEST DESCENT METHOD FOR QUASICONVEX FUNCTION
RIEMANNIAN MANIFOLDS**

Elsa Marisa Quispe Cárdenas

Enero - 2008

Adviser: DSc. Erik Alex Papa Quiroz.

Obtained Degree: Mathematician.

In this work, we proof the full convergence of the steepest descent method whit a generalized Armijo search to solve minimization problems whit quasiconvex objetive functions defined on complete riemanniana manifolds whit nonnegative sectional curvature. Previous convergence results obtained in euclidian spaces are particular case of our approach. Moreover, we introduce a class of diagonal metrics on \mathbb{R}_{++}^n and study its geometrical properties as: geodesics, sectional curvature, riemannian distances.

Keyboards:

Riemannian manifolds.

Steepest descent method.

Quasiconvex functions.

Full convergence.

Índice

Introducción	1
1 Preliminares	4
1.1 Símbolos y Notaciones	4
1.2 Definiciones Básicas	5
2 Elementos de Geometría Riemanniana	8
2.1 Introducción	8
2.2 Variedades diferenciables	9
2.3 Aplicaciones diferenciables entre variedades	13
2.4 Espacio tangente a una variedad diferenciable	15
2.5 Métricas riemannianas en variedades diferenciables	18
2.6 Campos de vectores, conexiones afines y deriva da covariante	23
2.7 Curvatura de una variedad riemanniana	35
2.8 Gradiente y Hessiana en una variedad riemanniana	39
2.9 Variedades completas	43
3 El Problema de Optimización y sus Condiciones de Optimalidad	47
3.1 Existencia de puntos de mínimo global	48
3.2 Caracterización de puntos de mínimo local	49
3.3 Elementos del análisis convexo	51
3.3.1 Convexidad en una variedad riemanniana	52
3.3.2 Funciones cuasi-convexas y pseudoconvexas	56

4 Método del Máximo Descenso	58
4.1 Convergencia para el caso cuasi-convexo	60
4.1.1 Método con búsqueda de Armijo generalizado	61
Bibliografía	69

Introducción

Fué N. I. Lobachevski quien se atrevió a darnos la posibilidad de la existencia de una geometría no euclidiana [1], proponiendo nuevas ideas, relacionando la geometría con la realidad material, el método, el alcance y sus aplicaciones. A raíz de estas nuevas ideas, los matemáticos estudian diversos espacios, además del euclidiano, entre ellos los espacios de Lobachevski [1], los proyectivos, de infinitas dimensiones, los riemannianos, topológicos entre otros. La utilización de la geometría riemanniana [1] es una de las mayores aplicaciones que se observa en Física, recordemos que fué Einstein quién en 1915 aplicó las ideas de la geometría de Riemann a la teoría de la gravitación universal que origina consecuentemente las diferentes aplicaciones de este tipo de geometrías en diversos campos de la ciencia y en la actualidad de mucho éxito en *Optimización*, como iremos viendo. La relación entre los métodos de Optimización Matemática y la geometría riemanniana data por lo menos del año 1972, con el trabajo desarrollado por LUENBERGER[15], donde usando el método de descenso geodésico obtiene la tasa de convergencia del método del gradiente proyectado para el problema de $\min f(x)$, sujeto a $h(x) = 0$, donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$, $n > m$. Esta línea de investigación tuvo continuidad con GABAY[10] en el año de 1982 donde, del punto de vista de esta teoría, estudia el método de gradiente reducido, generaliza los métodos de Cuasi-Newton obteniendo convergencia superlineal. También hace un análisis computacional mostrando que la teoría y la práctica interrelacionadas pueden dar buenos resultados.

El método de máximo descenso, estudiado por Cauchy en 1847, es uno de los métodos más antiguos y conocidos en la literatura para resolver problemas de optimización con funciones objetivo continuamente diferenciables. Sin embargo para una función arbitraria los resultados de convergencia no son muy fuertes ya que la convergencia global,

como también la existencia de puntos de acumulación no son garantizados. Solamente podemos asegurar que cualquier punto de acumulación, si existe, es un punto crítico del problema. La situación es muy diferente cuando la función objetivo es convexa, porque asumiendo solamente que el conjunto de soluciones óptimas es no vacío, el método de máximo descenso con búsqueda de ARMIJO[2] y con una regularización proximal converge a un punto óptimo. Este método en variedades riemannianas, considerando un problema de optimización con función objetivo arbitraria, fue estudiado por UDRISTE[22], SMITH[21] y RAPCSÁK[20], obteniendo los mismos resultados clásicos de convergencia. Para el caso convexo en estas variedades con curvatura seccional no negativa, la convergencia global usando la regla de búsqueda de Armijo, pasos fijos y una regularización proximal, fue generalizada por da CRUZ NETO et. al.[6] y para funciones cuasi-convexas solamente en espacios euclidianos, por KIWIEL y MURTY[12]. Esta tesis, esta orientada a extender los resultados de Kiwiel y Murty en variedades riemannianas completas y con curvatura seccional no negativa.

En el Capítulo 1, resumimos algunos símbolos y notaciones previas al desarrollo de conceptos que iremos usando en el todo el trabajo. En el Capítulo 2, presentamos elementos básicos de la geometría riemanniana BOOTHBY[3], de CARMO[7], [8], LAGES[13],[14] y su relación con la Optimización OLIVEIRA[17], damos ejemplos de las métricas más conocidas y estudiamos una clase particular de métricas riemannianas diagonales, definidas en el ortante positivo \mathbb{R}_{++}^n y el hipercubo abierto $(0, 1)^n$, espacios naturales donde se definen los problemas de optimización, obteniendo propiedades geométricas importantes como curvatura cero, ecuaciones sencillas para hallar geodésicas y condiciones suficientes para garantizar que la variedad riemanniana sea completa. En el Capítulo 3 presentamos el problema de optimización sobre una variedad riemanniana y desarrollamos sus condiciones de optimalidad, caracterizamos los puntos de mínimo, luego estudiamos la clase de funciones convexas y cuasi-convexas. Finalmente en el Capítulo 4, desarrollamos el método de máximo descenso, y analizamos la convergencia del método para resolver el problema de minimización usando funciones objetivo cuasi-convexas. Probamos que la sucesión generada por el método, usando la regla de búsqueda generalizada de Armijo, converge a un punto crítico de la función, mostrando la importancia del uso de métricas

diagonales que faciliten el análisis de un problema de optimización en variedades riemannianas. Nuestro resultado extiende la convergencia del método, del caso convexo para el caso cuasi-convexo generando el artículo [19] publicado por la revista: *Journal of Mathematical Analysis and Applications* (USA).

Capítulo 1

Preliminares

En este capítulo haremos un resumen de los resultados básicos necesarios al desarrollo de los subsiguientes capítulos. Las demostraciones serán, en principio, todas ellas referenciadas.

1.1 Símbolos y Notaciones

A lo largo de este trabajo, usaremos la siguiente simbología:

$$(0, 1)^n = (0, 1) \times (0, 1) \times \dots \times (0, 1).$$

$$\mathbb{R}_+^n = \{x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n : x_i > 0, i = 1, 2, \dots, n\}.$$

$$\mathbb{R}_{++}^n = \{x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n : x_i \geq 0, i = 1, 2, \dots, n\}.$$

Dados $x, y \in \mathbb{R}^n$, $(x, y) = \sum_{i=1}^n x_i y_i$: producto interno euclideo en \mathbb{R}^n .

$C^p(\Omega) = \{f : \Omega \rightarrow \mathbb{R} : f \text{ es diferenciable de orden } p\}$ es el conjunto de funciones p veces diferenciables en un dominio abierto Ω . Si $p = \infty$, entonces $C^\infty(\Omega)$ es el conjunto de funciones infinitamente diferenciables.

M : variedad diferenciable.

$T_p M$: es el espacio tangente a M en el punto p .

\mathcal{H} : es el conjunto de campos de vectores $X \in T_p M$.

∇ : es la conexión afín del conjunto de campo de vectores \mathcal{H} .

$X(p)$: es un campo vectorial aplicado en el punto p .

$\text{grad } f(x)$: es el gradiente de f en el sentido de la derivada covariante.

H^f : es la matriz Hessiana de f .

1.2 Definiciones Básicas

Definición 1.2.1 Sea $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ una función diferenciable definida en un abierto U . Definimos la diferencial de F en el punto $q \in U$ como una aplicación

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

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

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

Se puede probar fácilmente (ver de CARMO[7] pp. 127-128) que la diferencial es una aplicación lineal que no depende de la curva α tal que $\alpha(0) = q$, además, si:

$$F(x) = F(x_1, x_2, \dots, x_n) = (F_1(x), F_2(x), \dots, F_n(x)),$$

la diferencial en el punto q , en las bases canónicas es:

$$dF_q = \begin{bmatrix} \frac{\partial F_1}{\partial x_1}(q) & \frac{\partial F_1}{\partial x_2}(q) & \dots & \frac{\partial F_1}{\partial x_n}(q) \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_m}{\partial x_1}(q) & \frac{\partial F_m}{\partial x_2}(q) & \dots & \frac{\partial F_m}{\partial x_n}(q) \end{bmatrix}.$$

Definición 1.2.2 Sea $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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 imágen $F(p)$, 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 $x \in F^{-1}(a)$.

Por un resultado de álgebra lineal obtenemos una equivalencia para la sobreyectividad de la diferencial dF_x :

Para todo $x \in F^{-1}(a)$, dF_x es sobreyectiva si y solamente si, el rango $(dF_x) = m \leq n$. Así: $a \in F(U)$ es valor regular si, y solamente si, el rango $(dF_x) = m$, para todo $x \in F^{-1}(a)$.

En particular si $m = 1$ tal que $a \in F(U)$ es valor regular si, y solamente si, $\nabla F(x) \neq 0$ para todo $x \in F^{-1}(a)$.

Definición 1.2.3 Dado un conjunto M , un subconjunto Γ de partes de M , se dice una topología en M si:

1. $\emptyset \in \Gamma$ y $M \in \Gamma$;
2. Si $A, B \in \Gamma$ entonces $A \cap B \in \Gamma$;
3. Si $(A_i)_{i \in I}$ es una familia de elementos de Γ , entonces $\bigcup_{i \in I} A_i \in \Gamma$.

El par (M, Γ) se dice espacio topológico, los elementos de Γ son llamados abiertos del espacio topológico (Γ, M) .

Definición 1.2.4 (Espacios de Hausdorff). Sea M un espacio topológico, diremos que M es un espacio de Hausdorff, si para cualquier par de elementos distintos en M , se tienen abiertos disjuntos de dichos elementos.

Definición 1.2.5 Sea M un conjunto, una métrica es una aplicación $d : M \times M \rightarrow \mathbb{R}$ tal que para todo $x, y \in M$ se satisfacen las siguientes condiciones:

$$d_1 : d(x, y) \geq 0, d(x, y) = 0 \text{ si y sólo si } x = y;$$

$$d_2 : d(x, y) = d(y, x);$$

$$d_3 : d(x, z) \leq d(x, y) + d(y, z).$$

El par (M, d) se llama espacio métrico.

Definición 1.2.6 Una sucesión $\{x^m\}$ en un espacio métrico M se llama de Cauchy cuando para todo $\epsilon > 0$ dado, existe $n_0 \in \mathbb{N}$ tal que para todo $m, n > n_0$, se tiene $d(x^m, x^n) < \epsilon$.

Definición 1.2.7 El espacio métrico (M, d) es completo cuando toda sucesión de Cauchy en M es convergente.

Definición 1.2.8 Sea (X, d) un espacio métrico completo. Una sucesión $\{y^k\}$, $k \geq 0$, de X es cuasi-Fejér convergente al conjunto $U \subset X$, si para cada $u \in U$ existe una sucesión $\{\epsilon_k\} \subseteq \mathbb{R}$ tal que $\epsilon_k \geq 0$, $\sum_{k=0}^{+\infty} \epsilon_k < +\infty$ y $d^2(y^{k+1}, u) \leq d^2(y^k, u) + \epsilon_k$.

Teorema 1.2.1 *En un espacio métrico completo (X, d) , si $\{y^k\}$ es cuasi-Fejér convergente para un conjunto $U \subseteq X$, entonces $\{y^k\}$ es limitada. Si además, un punto de acumulación \bar{y} de $\{y^k\}$ pertenece a U . Entonces $\{y^k\}$ converge y $\lim_{k \rightarrow \infty} y^k = \bar{y}$.*

Demostración. Análogo a BURACHIK[4]. ■

Capítulo 2

Elementos de Geometría Riemanniana

2.1 Introducción

Las nociones de geometría riemanniana fueron introducidas por G. Riemann un 10 de Junio de 1854 a través de una disertación titulada: Sobre las hipótesis que están en los fundamentos de la geometría. En él afirma que toda colección continua de fenómenos homogéneos puede considerarse como un espacio. Estas ideas dieron origen a lo que hoy conocemos como geometría riemanniana, para ser más precisos, el tópico de la geometría diferencial que consiste en el estudio de las variedades diferenciales y si en estas variedades se introduce una particular forma de medir longitudes de curva, entonces estaremos hablando de variedades riemannianas, en esta geometría confluyen las ramas de la matemática, como son el álgebra, análisis, topología y la geometría, su abordaje dependera del problema a estudiarse. Es así que, en éste capítulo presentamos los conceptos principales de estas ideas que usaremos a lo largo de esta tesis, como son: variedades diferenciables, aplicaciones diferenciables entre variedades diferenciables y los espacios tangentes a estas variedades, también definiremos métrica riemanniana, geodésica, curvatura, gradiente y Hessiano de funciones o matriz Hessiana de una función, en una variedad riemanniana. En lo que concierne a métricas, presentamos una clase de métricas riemannianas diagonales, herramientas que nos permiten obtener propiedades interesantes para desarrollar nuevos algoritmos

en Optimización, esto es, que sobre una variedad riemanniana se puede derivar un campo vectorial tangente a lo largo de una curva, a través de la llamada derivada covariante a lo largo de curvas que depende de la métrica. Formalizaremos estos resultados en las siguientes secciones.

2.2 Variedades diferenciables

Una variedad diferenciable, a grosso modo, es un espacio topológico (no necesariamente vectorial) semejante localmente al espacio euclidiano \mathbb{R}^n cuya relación tiene el soporte de la diferenciabilidad. En esta sección presentamos estas ideas y daremos algunos ejemplos de variedades diferenciables relacionadas con problemas de Optimización. Para una demostración rigurosa de los resultados aquí presentados, referenciamos a Manfredo do CARMO[8], ELON LAGES[13] y [14], BOOTHBY[3], HICKS[11]. En todo este capítulo, el término diferenciable de una función o aplicación significará que es infinitamente diferenciable.

Definición 2.2.1 (*Superficie regular de \mathbb{R}^n*). Un subconjunto $S \subset \mathbb{R}^n$, es una superficie regular de \mathbb{R}^n de dimensión $k \leq n$ si para cada $p \in S$ existe una vecindad V de p en \mathbb{R}^n , un subconjunto abierto $U \subset \mathbb{R}^k$ y una aplicación sobreyectiva $\mathcal{X} : U \rightarrow S \cap V$ tal que:

1. \mathcal{X} es diferenciable en U .
2. \mathcal{X} es homeomorfismo.
3. Para todo $q \in U$, $d\mathcal{X}_q : \mathbb{R}^k \rightarrow \mathbb{R}^n$ es inyectiva, donde $d\mathcal{X}_q$ es la diferencial de \mathcal{X} en el punto q .

Véase un gráfico de la definición de superficie regular en \mathbb{R}^3 .

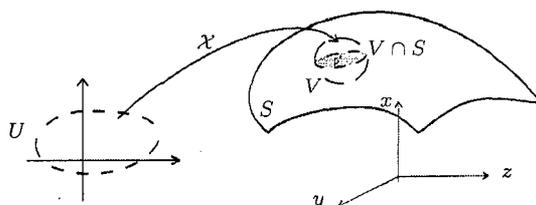


Figura 1. Superficie regular

Para cada $p \in S$, la aplicación $\mathcal{X} : U \rightarrow V \cap S$ es llamada **parametrización de S en p** , o **sistema de coordenadas locales en p** . $V \cap S$ es llamada vecindad coordinada de p .

Proposición 2.2.1 *Sea U un subconjunto abierto de \mathbb{R}^n y $F : U \rightarrow \mathbb{R}^m$ una función diferenciable en U con valor regular $a \in \mathbb{R}^m$, entonces $F^{-1}(a)$ es una superficie regular de dimensión $n - m$.*

Corolario 2.2.1 *Sea U un subconjunto abierto de \mathbb{R}^n y $f : U \rightarrow \mathbb{R}$ una función diferenciable tal que $(\nabla f)(x) \neq 0$, para todo $x \in f^{-1}(a)$. Entonces $S = f^{-1}(a)$ es una superficie regular.*

La Proposición 2.2.1 permite mostrar una familia de superficies regulares útiles en el contexto de la Optimización Matemática.

Ejemplo 2.2.1 *Consideremos los problemas de Optimización Lineal:*

$$\begin{array}{ll} (P) \min & c^T x \\ & \text{s.a. } Ax = b \\ & x \geq 0. \end{array} \qquad \begin{array}{ll} (D) \max & b^T \lambda \\ & \text{s.a. } A^T \lambda + s = c \\ & s \geq 0. \end{array}$$

donde: $x, s, c \in \mathbb{R}^n$; $\lambda, b \in \mathbb{R}^m$ y $A \in \mathbb{R}^{m \times n}$ es de rango $m < n$. El problema (P) es llamado primal y (D) el dual de (P).

1. **Restricciones estrictas primales.** Si $S = \{x \in \mathbb{R}_{++}^n : Ax = b\}$ es el conjunto de las restricciones estrictas del problema (P), definiendo la función $F : \mathbb{R}_{++}^n \rightarrow \mathbb{R}^m$, por $F(x) = Ax - b$ se tiene que $dF_x = A$, para todo $x \in \mathbb{R}_{++}^n$. Por tener la matriz A rango m y aplicando la Proposición 2.2.1 tenemos que $F^{-1}(0) = S$ es una superficie regular de dimensión $n - m$.
2. **Restricciones estrictas duales.** Si $S = \{(\lambda, s) \in \mathbb{R}^m \times \mathbb{R}_{++}^n : A^T \lambda + s = c\}$ es el conjunto de las restricciones estrictas del problema (D), definiendo la función $F : \mathbb{R}^m \times \mathbb{R}_{++}^n \rightarrow \mathbb{R}^n$, por:

$$F(\lambda, s) = A^T \lambda + s - c = [A^T \ I] \begin{bmatrix} \lambda \\ s \end{bmatrix} - c$$

Se tiene $dF_{(\lambda,s)} = [A^T \ I]$ con rango n , para todo $(\lambda, s) \in \mathbb{R}^m \times \mathbb{R}_{++}^n$. Aplicando la Proposición 2.2.1, $F^{-1}(0) = S$ es una superficie regular de dimensión m .

Ejemplo 2.2.2 Consideremos el problema en Optimización no Lineal:

$$\begin{aligned} \min \quad & f(x) \\ \text{s.a} \quad & h(x) = 0 \\ & x \in U \end{aligned}$$

donde U es un abierto de \mathbb{R}^n , $f : U \rightarrow \mathbb{R}$ y $h : U \rightarrow \mathbb{R}^m$ son funciones dadas.

Si h es diferenciable y su matriz Jacobiana en el punto x , $J_h(x)$ tiene rango m entonces, el conjunto $\{x \in U : h(x) = 0\}$ es una superficie regular. Como casos particulares tenemos que los conjuntos $\{x \in \mathbb{R}^n : h(x) = 0\}$ y $\{x \in \mathbb{R}^n : h(x) = 0 \text{ y } x > 0\}$ son superficies regulares.

Una propiedad importante de las superficies regulares, es que ella no depende del cambio de parámetros, es decir, cualquier otra parametrización de la superficie en un punto $p \in S$ sigue manteniendo las propiedades diferenciables. Este resultado nos servirá para generalizar la definición de superficie regular a variedad diferenciable.

Definición 2.2.2 (Cambio de parámetros). Sean $\mathcal{X} \rightarrow S$ y $\mathcal{Y} \rightarrow S$ dos parametrizaciones de S en el punto p tales que

$$W = \mathcal{X}(U) \cap \mathcal{Y}(V) \neq \emptyset.$$

La aplicación $\mathcal{Y}^{-1} \circ \mathcal{X} : \mathcal{X}^{-1}(W) \rightarrow \mathcal{Y}^{-1}(W)$ es llamada cambio de parámetros.

Proposición 2.2.2 Sea S una superficie regular de \mathbb{R}^n de dimensión k . El cambio de parámetros $\mathcal{Y}^{-1} \circ \mathcal{X} : \mathcal{X}^{-1}(W) \rightarrow \mathcal{Y}^{-1}(W)$ es un difeomorfismo.

La noción de variedad diferenciable que definimos a seguir es necesaria para poder extender los métodos del cálculo diferencial a espacios más generales. Como veremos posteriormente, una superficie regular será un claro ejemplo de variedad diferenciable.

Definición 2.2.3 (*Variedad diferenciable*). Una variedad diferenciable de dimensión n es un conjunto M y una familia de aplicaciones inyectivas $\mathcal{X}_\alpha : U_\alpha \rightarrow M$, $\alpha \in I$ (conjunto de parámetros), definidos en abiertos U_α de \mathbb{R}^n en M tales que se cumplen las siguientes condiciones:

1. $M = \bigcup_{\alpha \in I} \mathcal{X}_\alpha(U_\alpha)$.
2. Para todo par $\mathcal{X}_\alpha, \mathcal{X}_\beta$ 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**. Una familia $\{(U_\alpha, \mathcal{X}_\alpha)\}$ satisfaciendo los items 1 y 2 es llamada **estructura diferenciable** de M . Así, la variedad es un conjunto M con una estructura diferenciable.

Una estructura diferenciable en una variedad diferenciable M induce de forma natural una topología en M definido por:

$A \subset M$ es abierto en M si para todo $\alpha \in I$, $\mathcal{X}_\alpha^{-1}(A \cap \mathcal{X}_\alpha(U_\alpha))$ es abierto en \mathbb{R}^n .

Observemos que la topología es definida de tal modo que los conjuntos $\mathcal{X}_\alpha(U_\alpha)$ son abiertos y las aplicaciones \mathcal{X}_α son continuas.

Debido a la Proposición 2.2.2 podemos enunciar, el siguiente resultado.

Proposición 2.2.3 *Toda superficie regular de \mathbb{R}^n de dimensión k es una variedad diferenciable de la misma dimensión.*

Definición 2.2.4 (*Variedad de Hausdorff de base numerable*). Una variedad diferenciable M , es llamada *variedad de Hausdorff* si, M con la topología dada, es un espacio de Hausdorff.

La variedad diferenciable M tiene base numerable si ella puede ser cubierta por una cantidad numerable de vecindades coordenadas, esto es, si existe una sucesión $\{\mathcal{X}_n(U_n)\}$, $n \in \mathbb{N}$, de vecindades coordenadas tal que: $M = \bigcup_{n \in \mathbb{N}} \mathcal{X}_n(U_n)$.

En todo este capítulo asumiremos que la variedad diferenciable M es de Hausdorff y de base numerable.

2.3 Aplicaciones diferenciables entre variedades

Definición 2.3.1 Sea $f : U \subset M \rightarrow \mathbb{R}$, donde U es un subconjunto abierto de la 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 función compuesta $f \circ \mathcal{X}_\alpha : U_\alpha \subset \mathbb{R}^n \rightarrow \mathbb{R}$ es diferenciable en $\mathcal{X}_\alpha^{-1}(p)$. Se dice que f es diferenciable en U si es diferenciable en todo punto de U .

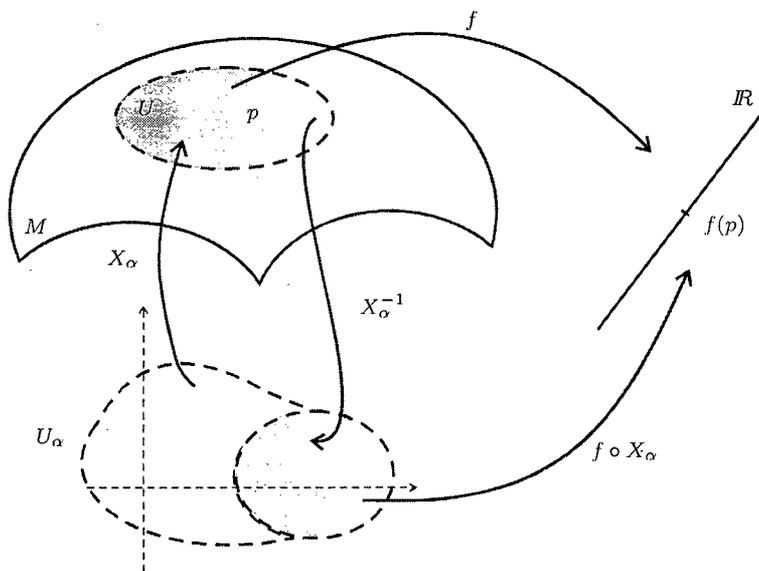


Figura 2. f es diferenciable en U .

Una consecuencia inmediata del ítem 2 de la Definición 2.2.3 es que, la diferenciability de una función de valores reales definida sobre una variedad diferenciable M no depende de la elección de la parametrización. En efecto, sea $\mathcal{X}_\beta : U_\beta \subset \mathbb{R}^n \rightarrow M$, otra parametrización tal que, $p \in \mathcal{X}_\beta(U_\beta) \subset U$. Podemos expresar:

$$f \circ \mathcal{X}_\beta = f \circ \mathcal{X}_\alpha \circ \mathcal{X}_\alpha^{-1} \circ \mathcal{X}_\beta : U_\beta \subset \mathbb{R}^n \rightarrow \mathbb{R}.$$

Como $f \circ \mathcal{X}_\alpha$ es diferenciable por definición y $\mathcal{X}_\alpha^{-1} \circ \mathcal{X}_\beta$ es diferenciable por ser cambio de parámetros, entonces $f \circ \mathcal{X}_\beta$ es también diferenciable.

Definición 2.3.2 Una curva sobre una variedad diferenciable M es una función $\gamma : I \rightarrow M$ donde $I = (-\varepsilon, \varepsilon)$. Diremos que γ 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 función compuesta $\beta = \mathcal{X}_\alpha^{-1} \circ \gamma : I \rightarrow U_\alpha$ es diferenciable en t_0 , donde $\gamma(I) \subset \mathcal{X}_\alpha(U_\alpha)$. Si γ es diferenciable en todo $t \in I$, diremos que γ es diferenciable en I .

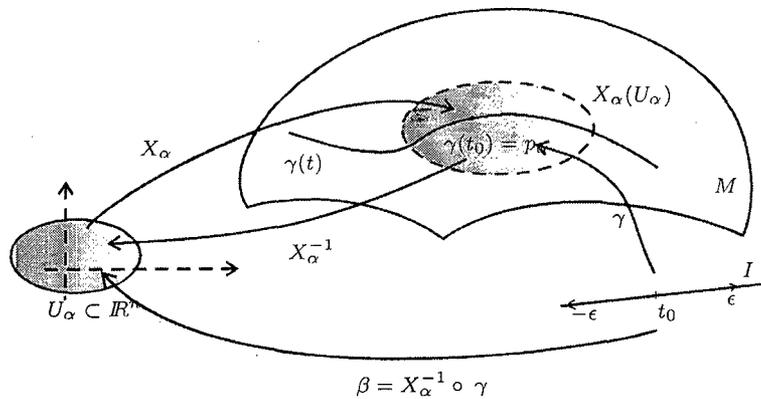


Figura 3. γ es diferenciable en $t \in I$.

La definición de diferenciabilidad puede ser extendida para aplicaciones entre variedades.

Definición 2.3.3 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 dados:

$$\mathcal{X}_1 : U_1 \subset \mathbb{R}^n \rightarrow M_1$$

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 $\mathcal{X}_1^{-1}(p)$. Esta última aplicación es llamada expresión de φ en las parametrizaciones \mathcal{X}_1 y \mathcal{X}_2 . φ es diferenciable en un abierto de M_1 si es diferenciable en todos los puntos del abierto.

Análogamente al caso de funciones de valores reales, se muestra que la definición no depende de las parametrizaciones elejidas.

Observación 2.3.1 Una consecuencia de la Definición 2.3.3 es que, si $\mathcal{X} : U \rightarrow M$ es una parametrización de M en el punto p entonces $\mathcal{X}^{-1} : \mathcal{X}(U) \subset M \rightarrow \mathbb{R}^n$ es diferenciable.

Definición 2.3.4 (Difeomorfismo entre variedades diferenciables). Sea $\varphi : M_1 \rightarrow M_2$ una aplicación diferenciable entre dos variedades diferenciables. Decimos que φ

es difeomorfismo si φ es biyectiva y φ^{-1} es diferenciable. φ es difeomorfismo local en $p \in M_1$, si existen vecindades U de p y V de $\varphi(p)$ tal que $\varphi : U \rightarrow V$ es difeomorfismo.

Observación 2.3.2 De la observación 2.3.1, concluimos que cualquier parametrización $\mathcal{X} : U \subset \mathbb{R}^n \rightarrow \mathcal{X}(U) \subset M$, es un difeomorfismo. Por esta razón, muchas veces para facilitar la notación se identifica $\mathcal{X}(U) \equiv U$.

2.4 Espacio tangente a una variedad diferenciable

Las consideraciones a seguir motivan la definición que extiende a variedades diferenciables la noción de vector tangente. Para superficies de \mathbb{R}^3 , un vector tangente en un punto p de la superficie es definida como el “vector velocidad” en \mathbb{R}^3 de una curva de la superficie pasando por p . Como en variedades diferenciables no disponemos del soporte de un espacio ambiente, precisamos de una propiedad característica del vector tangente que substituya la noción de velocidad.

Recordemos algunas formalidades en superficies regulares.

Sea $\varepsilon > 0$ suficientemente pequeño y una curva $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^n$ tal que:

$$\gamma(t) = (\gamma_1(t), \dots, \gamma_n(t)),$$

con $\gamma(0) = p$ y $\gamma'(0) = (\gamma'_1(0), \dots, \gamma'_n(0)) = v \in \mathbb{R}^n$. Sea además una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferenciable definida en una vecindad de p . Podemos restringir f a la curva γ y calcular la **derivada direccional** de f en la dirección de $v \in \mathbb{R}^n$:

$$\left. \frac{d(f \circ \gamma)}{dt}(t) \right|_{t=0} = \sum_{i=1}^n \frac{\partial f}{\partial \gamma_i}(\gamma(0)) \frac{d\gamma_i}{dt}(0) = \left(\sum_{i=1}^n \gamma'_i(0) \left(\frac{\partial}{\partial \gamma_i} \right)_p \right) f.$$

Por tanto la derivada direccional en la dirección de v es un operador sobre funciones diferenciables que depende unicamente de v y esta es la propiedad característica que usaremos para definir un vector tangente en variedades.

Definición 2.4.1 (Vector tangente en un punto de una variedad diferenciable). Sea M una variedad diferenciable. Consideremos una curva diferenciable

$$\gamma : (-\varepsilon, \varepsilon) \rightarrow M,$$

donde $\gamma(0) = p$ y sea $D_p = \{f : M \rightarrow \mathbb{R} : f \text{ es diferenciable en } p\}$. Definimos el vector tangente a la curva γ en $t = 0$ como la función $\gamma'(0) : D_p \rightarrow \mathbb{R}$ dada por:

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

Un vector tangente en p es el vector tangente en $t = 0$ de alguna curva $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ con $\gamma(0) = p$.

Si M es una superficie regular de dimensión $k \leq n$, esto es $M \subset \mathbb{R}^n$, definimos el vector tangente en el punto p como el vector velocidad en \mathbb{R}^n , esto es,

$$\gamma'(0) = (\gamma'_1(0), \gamma'_2(0), \dots, \gamma'_n(0)).$$

Definición 2.4.2 (Espacio tangente a una variedad diferenciable). 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 2.4.1 Si para una parametrización $\mathcal{X} : U \subset \mathbb{R}^n \rightarrow M$ con $p = \mathcal{X}(0)$ y $q \in U$, podemos restringir la función $f \in D_p$ y la curva $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ en esta parametrización:

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

(identificación: $f \circ \mathcal{X} \equiv f$).

Podemos escribir también,

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

Por definición tenemos:

$$\begin{aligned} \gamma'(0)f &= \left. \frac{d(f \circ \gamma)}{dt}(t) \right|_{t=0} = \left. \frac{d(f \circ \mathcal{X} \circ \mathcal{X}^{-1} \circ \gamma)}{dt}(t) \right|_{t=0} \\ &= \left. \frac{d}{dt}(f(q_1(t), q_2(t), \dots, q_n(t))) \right|_{t=0} \end{aligned}$$

entonces:

$$\gamma'(0)f = \sum_{i=1}^n q'_i(0) \cdot \left(\frac{\partial f}{\partial q_i}(p) \right) = \left(\sum_{i=1}^n q'_i(0) \cdot \left(\frac{\partial}{\partial q_i} \right)_0 \right) f.$$

Así,

$$\gamma'(0) = \sum_{i=1}^n q_i'(0) \cdot \left(\frac{\partial}{\partial q_i} \right)_0 \quad (2.1)$$

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

Observación 2.4.2 Para una curva coordenada en U , $\beta_i(x_i) = (0, \dots, x_i, \dots, 0)$, se tiene que la composición $\mathcal{X} \circ \beta_i = \gamma_i$ es una curva coordenada sobre M y de la ecuación anterior, $\gamma_i'(0) = \left(\frac{\partial}{\partial x_i} \right)_0$. Se sigue que $\left(\frac{\partial}{\partial x_i} \right)_0$ es el vector tangente a la curva coordenada $\gamma_i(t)$.

Observación 2.4.3 De la elección de una parametrización obtenemos n “vectores” $\left(\left(\frac{\partial}{\partial x_i} \right)_p, i = 1, \dots, n \right)$ en $T_p M$ que generan, por (2.1), los vectores en $T_p M$.

Observación 2.4.4 Sea M una variedad diferenciable, el fibrado tangente de M es definido por:

$$TM = \{(p, v); p \in M / v \in T_p M\}.$$

TM puede ser unido de una estructura diferenciable transformándose así en una variedad diferenciable (ver de CARMO[7], pag. 15 para su demostración).

En los siguientes resultados presentamos ejemplos de espacios tangentes.

Proposición 2.4.1 El espacio tangente de una variedad diferenciable que es un subconjunto abierto de \mathbb{R}^n es el propio \mathbb{R}^n .

Como consecuencia de esta proposición se tiene:

a). $T_p \mathbb{R}^n = \mathbb{R}^n, T_p \mathbb{R}_{++}^n = \mathbb{R}^n$.

b). Si $M = \{(\lambda, s) \in \mathbb{R}^m \times \mathbb{R}^n : s > 0\}$, entonces $T_p M = \mathbb{R}^{m+n}$.

Proposición 2.4.2 Sea $M = F^{-1}(a)$ una variedad de dimensión $n - m$, donde la aplicación $F : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ es una función diferenciable, U es abierto y a es un valor regular de F , entonces: $T_p M = T_p(F^{-1}(a)) = \text{Ker}(dF_p)$.

Ejemplo 2.4.1 Si $M = \{x \in \mathbb{R}_{++}^n : Ax = b\}$, donde $A \in \mathbb{R}^{m \times n}$ tiene rango $m < n$, entonces:

$$T_p M = \text{Ker} A = \{\Delta x \in \mathbb{R}^n : A \Delta x = 0\}.$$

En efecto, la función que define M es $F : \mathbb{R}_{++}^n \rightarrow \mathbb{R}^m$ tal que $F(x) = Ax - b$, la diferencial de F en el punto $p \in M$ es $dF_p = A$, luego aplicando la Proposición 2.4.2 obtenemos el resultado.

Ejemplo 2.4.2 Sea $h : U \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ una función diferenciable con Jacobiano $J_h(x) = dh_x$ con rango m . Consideremos la variedad $M = h^{-1}(0) = \{x \in U, h(x) = 0\}$, entonces:

$$T_p(h^{-1}(0)) = \text{Ker}(J_h(x)).$$

Corolario 2.4.1 Sea $D \subset \mathbb{R}^n$ un conjunto abierto y $f : D \rightarrow \mathbb{R}$ una función diferenciable tal que $\nabla f(x) \neq 0$, para todo $x \in f^{-1}(a)$ vemos que, $M = f^{-1}(a)$ es una variedad diferenciable. Entonces para cada $p \in M$,

$$T_p(f^{-1}(a)) = \nabla f(p)^\perp.$$

Proposición 2.4.3 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 $\alpha : (-\epsilon, \epsilon) \rightarrow M$ con $\alpha(0) = p$, $\alpha'(0) = v$. 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 de α . Esta aplicación es llamada la diferencial de φ en p .

Proposición 2.4.4 Sea M_1 y M_2 dos variedades diferenciables. Si $\varphi : M_1 \rightarrow M_2$ es un difeomorfismo, entonces $d\varphi_p : T_p M_1 \rightarrow T_{\varphi(p)} M_2$ es un isomorfismo.

2.5 Métricas riemannianas en variedades diferenciables

Las métricas en un espacio son muy importantes porque nos permiten medir distancias, calcular errores, longitudes de curvas, etc. Cuando tenemos una curva

parametrizada en \mathbb{R}^n , $\gamma(t) = (\gamma_1(t), \gamma_2(t), \dots, \gamma_n(t))$ donde t pertenece a algun intervalo I de \mathbb{R} , la longitud de arco de la curva generada por $\gamma(t)$ es medida por:

$$\ell(\gamma) = \int_I \|v(t)\| dt$$

donde $v(t) = (\gamma_1'(t), \gamma_2'(t), \dots, \gamma_n'(t))$ y $\|\cdot\|$ representa la norma euclidea. Así, la longitud de la curva depende de la norma del vector velocidad definido por la métrica usual en \mathbb{R}^n . Ahora, si nuestro espacio es una variedad diferenciable M y tenemos definida una curva en ella, entonces la longitud de arco de la curva será obtenida por la **medida** realizada en el vector perteneciente al espacio tangente en cada punto. Necesitamos entonces definir una métrica en el espacio tangente $T_p M$ para cada $p \in M$. Recordemos que además del producto interno clásico:

$$(v, w)_p = \sum_i v_i w_i,$$

podemos definir otro producto interno:

$$\langle v, w \rangle_p = \sum_{i,j} g_{ij} v_i w_j = (Gv, w),$$

donde $G = (g_{ij})$ es una matriz simétrica definida positiva. Esta definición aparece de modo natural al realizar un cambio de coordenadas. En efecto, sean $x = (x_1, x_2, \dots, x_n)$ y $z = (z_1, z_2, \dots, z_n)$ tal que $x(t) = x(z(t))$, esto es,

$$x(t) = (x_1(z_1(t), z_2(t), \dots, z_n(t)), x_2(z_1(t), z_2(t), \dots, z_n(t)), \dots, x_n(z_1(t), z_2(t), \dots, z_n(t))),$$

entonces:

$$\frac{dx_i}{dt} = \sum_{j=1}^n \left(\frac{\partial x_i}{\partial z_j} \frac{\partial z_j}{\partial t} \right), \text{ para todo } i = 1, 2, \dots, n.$$

Denotando $v^x = (v_1^x, v_2^x, \dots, v_n^x)$ y $v^z = (v_1^z, v_2^z, \dots, v_n^z)$, donde $v_i^x = \frac{dx_i}{dt}$ y $v_i^z = \frac{dz_i}{dt}$, tenemos:

$$\|v^x\|^2 = (v^x, v^x) = \sum_{i=1}^n \left(\frac{dx_i}{dt} \right)^2.$$

Como:

$$\left(\frac{dx_i}{dt} \right)^2 = \left(\sum_{j=1}^n \frac{\partial x_i}{\partial z_j} v_j^z \right)^2 = \sum_{j=1}^n \left(\frac{\partial x_i}{\partial z_j} v_j^z \sum_{k=1}^n \frac{\partial x_i}{\partial z_k} v_k^z \right) = \sum_{j=1}^n \left(\sum_{k=1}^n \frac{\partial x_i}{\partial z_k} v_k^z \frac{\partial x_i}{\partial z_j} \right) v_j^z,$$

entonces:

$$\|v^x\|^2 = \sum_{i=1}^n \sum_{j=1}^n \left(\sum_{k=1}^n \frac{\partial x_i}{\partial z_j} \frac{\partial x_i}{\partial z_k} v_k^z \right) v_j^z,$$

conmutando sumandos:

$$\|v^x\|^2 = \sum_{j=1}^n \sum_{k=1}^n \left(\sum_{i=1}^n \frac{\partial x_i}{\partial z_j} \frac{\partial x_i}{\partial z_k} \right) v_k^z v_j^z.$$

Haciendo un cambio k por i e i por k obtenemos:

$$\|v^x\|^2 = \sum_{i,j=1}^n \left(\sum_{k=1}^n \frac{\partial x_k}{\partial z_j} \frac{\partial x_k}{\partial z_i} \right) v_i^z v_j^z.$$

Definiendo $g_{ij} = \left(\sum_{k=1}^n \frac{\partial x_k}{\partial z_j} \frac{\partial x_k}{\partial z_i} \right)$ se tiene finalmente que:

$$(v^x, v^x) = \|v^x\|^2 = \sum_{i,j=1}^n g_{ij} v_i^z v_j^z = (Gv^z, v^z).$$

Queda claro que un cambio en el sistema de coordenadas no altera las métricas. En efecto, si $v = G^{1\setminus 2}w$ tenemos que:

$$(v, v)_p = (G^{1\setminus 2}w, G^{1\setminus 2}w)_p = (Gw, w)_p = \langle w, w \rangle_p.$$

Métrica riemanniana.

Definición 2.5.1 Sea S una variedad diferenciable. Una métrica riemanniana es una aplicación que asocia a cada $p \in M$ un funcional $\langle \cdot, \cdot \rangle_p$

$$\langle \cdot, \cdot \rangle_p : T_p M \times T_p M \longrightarrow \mathbb{R},$$

de modo que se cumplen las siguientes condiciones:

1. $\langle \cdot, \cdot \rangle_p$ es un producto interno (bilineal, simétrica y definida positiva) para cada $p \in M$.
2. $\langle \cdot, \cdot \rangle_p$ varía diferenciablemente en el siguiente sentido: Si $\mathcal{X} : U \subset \mathbb{R}^n \longrightarrow M$ es un sistema de coordenadas en torno de p , con $\mathcal{X}(x_1, x_2, x_3, \dots, x_n) = q \in \mathcal{X}(U)$ y $\frac{\partial}{\partial x_i}(q) = d\mathcal{X}_q(0, 0, \dots, 0, 1, 0, \dots, 0, 0)$, entonces la función: $g_{ij} : U \longrightarrow \mathbb{R}$ 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 diferenciable.

Las funciones g_{ij} son llamadas expresiones de la métrica riemanniana en el sistema coordinado \mathcal{X} y la matriz $G = (g_{ij})$ es la representación de la métrica riemanniana. Como \mathcal{X} es un difeomorfismo (Observación 2.3.2) se tiene que $d\mathcal{X}_q : \mathbb{R}^n \rightarrow T_q M$ es un isomorfismo (Proposición 2.4.4) y así la matriz $G = (g_{ij})$ es invertible. Por tanto, toda métrica riemanniana tiene su matriz de representación invertible.

Definición 2.5.2 (*Variedad riemanniana*). Una variedad diferenciable para la cual se define una métrica riemanniana se denomina una **variedad riemanniana**.

Ejemplo 2.5.1 Sea $M = \mathbb{R}^n$, defina la parametrización $\mathcal{X} : \mathbb{R}^n \rightarrow \mathbb{R}^n$ tal que $\mathcal{X}(x_1, x_2, \dots, x_n) = (x_1, x_2, \dots, x_n)$.

Definamos la métrica:

$$\langle \cdot, \cdot \rangle_p : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}, \text{ definido por } \langle x, y \rangle_p = x^T y.$$

Sea $q \in \mathbb{R}^n$ entonces:

$$\frac{\partial}{\partial x_i}(q) = d\mathcal{X}_q e_i = e_i,$$

y así, $g_{ij} : U \rightarrow \mathbb{R}$ definidas por:

$$g_{ij}(x) = \left\langle \frac{\partial}{\partial x_i}(x), \frac{\partial}{\partial x_j}(x) \right\rangle_x = \langle e_i, e_j \rangle_x = e_i^T e_j = \delta_{ij},$$

son diferenciables en \mathbb{R}^n .

Luego $M = \mathbb{R}^n$, con la expresión de la métrica $G = Id$, es una variedad riemanniana, esto es, el espacio euclidiano es un ejemplo particular de variedad riemanniana.

Ejemplo 2.5.2 Sea M la variedad definida por el siguiente conjunto:

$$M = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 > 0\}.$$

Usamos la parametrización identidad, además $T_p M = \mathbb{R}^2$, con $p = (p_1, p_2) \in M$.

Definimos la aplicación:

$$\langle \cdot, \cdot \rangle_p : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$$

tal que:

$$\langle (x_1, x_2), (y_1, y_2) \rangle_p = \frac{1}{p_2}(x_1 y_1 + x_2 y_2).$$

Sea $q = (q_1, q_2) \in M$, entonces

$$\frac{\partial}{\partial x_i}(q) = d\mathcal{X}_q e_i = e_i, \text{ para todo } i = 1, 2.$$

Luego: $g_{ij} : M \rightarrow \mathbb{R}$ definidas por:

$$g_{ij}(x_1, x_2) = \frac{1}{x_2} \langle e_i, e_j \rangle = \frac{1}{x_2} \delta_{ij},$$

son diferenciables en M . Por lo tanto, M es una variedad riemanniana con la expresión de la métrica:

$$G(p) = \frac{1}{p_2} (Id)_{2 \times 2}.$$

Esta métrica es conocida como **métrica de Lobatchevsky o Poincaré**.

Los siguientes dos ejemplos son también variedades riemannianas para las métricas G según se definen, los cuales se demuestran bajo el mismo procedimiento que los dos últimos ejemplos anteriores.

Ejemplo 2.5.3 Si $M = \mathbb{R}_{++}^n$ y el funcional $\langle \cdot, \cdot \rangle_p : T_p \mathbb{R}_{++}^n \times T_p \mathbb{R}_{++}^n \rightarrow \mathbb{R}$ tal que:

$$\langle u, v \rangle_p = u^T G(p)v,$$

donde:

$$G(p) = \text{diag}(1/(h_i(p_i))^2),$$

entonces:

$$g_{ij}(x) = \frac{\delta_{ij}}{(h_2(x_i))^2}.$$

Ejemplo 2.5.4 Sea la variedad riemanniana $(\mathbb{R}^2, G(x))$, con $\langle u, v \rangle_p = u^T G(p)v$,

donde:

$$G(p) = \begin{bmatrix} 4p_1^2 + 1 & -2p_1 \\ -2p_1 & 1 \end{bmatrix}$$

es la métrica riemanniana dada por **Udriste**.

2.6 Campos de vectores, conexiones afines y derivada covariante

Introducimos los campos de vectores en los espacios tangentes a las variedades riemannianas. En Optimización Matemática ellos representan las direcciones, a partir de un punto dado, para algún algoritmo iterativo propuesto. Observando la trayectoria continua del algoritmo, esta tendrá sus características, como curvatura, que dependen obviamente de las características del campo. Surge así la necesidad de definir conceptos correspondientes al de derivada de funciones.

Definición 2.6.1 (*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:

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

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

Es útil pensar en campos vectoriales como aplicaciones $X : D \rightarrow F$ definidas por

$$(Xf)(p) = \sum_i a_i(p) \cdot \frac{\partial f}{\partial x_i}(p),$$

donde D es el conjunto de las funciones diferenciables sobre M y F es el conjunto de las funciones sobre M .

Como estamos interesados en trayectorias en M , consideraremos los campos restringidos a una curva.

Definición 2.6.2 (*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 $V(t) \in T_{\alpha(t)} M$. Se dice que V es diferenciable si para cada función diferenciable f en D , la función $V(t)f$ es una función diferenciable en I .

Sea X un campo definido en M , el campo X a lo largo de α será denotado $V(t) =$

$X(\alpha(t))$ y diremos que V es inducido por X .

El campo vectorial $d\mathcal{X}_{(\mathcal{X}^{-1} \circ \alpha)(t)}[(\mathcal{X}^{-1} \circ \alpha)'(t)] = \frac{d(\mathcal{X} \circ \mathcal{X}^{-1} \circ \alpha)(t)}{dt}$ denotado por $\frac{d\alpha}{dt}$, es llamado campo velocidad o tangente de α .

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 y $D = D(M) = \{f : M \rightarrow \mathbb{R} : f \in C^\infty\}$ el conjunto de funciones reales de clase C^∞ .

Definición 2.6.3 Una conexión afin 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 $f, g \in D$ verifique:

1. $\nabla_{(fX+gY)}Z = f\nabla_X Z + g\nabla_Y Z$;
2. $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$;
3. $\nabla_X fY = f\nabla_X Y + X(f)Y$, donde $X(f) = \sum_{i=1}^n a_i(\cdot) \frac{\partial f(\cdot)}{\partial x_i}$.

Considerando una curva diferenciable en M $\alpha : I \rightarrow M$, denotaremos el conjunto de campo de vectores a lo largo de esta curva como \mathcal{H}_α .

Proposición 2.6.1 Sea M una variedad diferenciable con una conexión afin ∇ . Entonces existe una única aplicación $\frac{D}{dt}$, donde a cada $V \in \mathcal{H}_\alpha$ se asocia otro campo en \mathcal{H}_α , denotado por $\frac{DV}{dt}$, tal que para todo $V, W \in \mathcal{H}_\alpha$ y $f : I \rightarrow \mathbb{R}$ una función diferenciable en I se cumplen:

- a. $\frac{D}{dt}(V + W) = \frac{DV}{dt} + \frac{DW}{dt}$.
- b. $\frac{D}{dt}(fV) = \frac{df}{dt}V + f\frac{DV}{dt}$.
- c. Si $V(t) = Y(\alpha(t))$, donde $Y \in \mathcal{H}$, entonces $\frac{DV}{dt} = \nabla_{\frac{d\alpha}{dt}} Y$.

$\frac{D}{dt}$ es llamada **Derivada Covariante**.

Observación 2.6.1 La Proposición 2.6.1 muestra que la elección de una conexión afin de M da origen a una única derivada covariante para cada campo vectorial a lo largo de una curva.

Observación 2.6.2 Al realizar la demostración de esta Proposición, encontramos una caracterización de la derivada covariante para una cierta parametrización \mathcal{X} , de acuerdo con:

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

Observación 2.6.3 La noción de conexión afin, ofrece una manera de derivar campo de vectores a lo largo de curvas. Así, en particular para el campo vectorial $V = \frac{d\alpha}{dt}$ tenemos:

$$\frac{D}{dt} \left(\frac{d\alpha}{dt} \right),$$

que llamaremos **aceleración de una curva α en M** .

Expresión de la conexión afin relativa a coordenadas locales.

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

$$X = \sum_{i=1}^n x_i \frac{\partial}{\partial x_i}, \quad Y = \sum_{i=1}^n y_i \frac{\partial}{\partial x_i},$$

donde $(\partial/\partial x_i)$ representan los vectores de la base del sistema de coordenadas locales.

Por simplicidad de notación expresaremos:

$$\frac{\partial}{\partial x_i} = X_i.$$

Entonces tenemos.

$$X = \sum_{i=1}^n x_i X_i, \quad Y = \sum_{i=1}^n y_i X_i.$$

Según las propiedades de la definición de la conexión afin:

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

Observe que $\nabla_{X_i} X_j \in \mathcal{H}$, pudiendo por tanto ser también representado através de una base local, esto es:

$$\nabla_{X_i} X_j = \sum_{k=1}^n \Gamma_{ij}^k X_k \tag{2.2}$$

que, substituyendo en la ecuación anterior, se obtiene:

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

Definición 2.6.4 (*Símbolos de Christoffel*). Los símbolos de Christoffel, o coeficientes de la conexión afín ∇ en U , son las funciones (diferenciables):

$$\Gamma_{ij}^k : U \subset M \rightarrow \mathbf{R}$$

definidas por (2.2).

Expresión de la derivada covariante en términos de 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 2.6.1 es:

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

y usando

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

tenemos

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

Así:

$$\frac{DV}{dt} = \sum_{k=1}^n \left(\frac{dv^k}{dt} + \sum_{i,j=1}^n v^j \frac{dx_i}{dt} \Gamma_{ij}^k \right) X_k \quad (2.3)$$

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

Geodésicas y Campos paralelos.

La derivada covariante permite definir el transporte paralelo a lo largo de curvas que dependen de la métrica, osea, que cambiando la métrica, cambia en general la manera de derivar campos vectoriales, en particular nos permite conocer geodésicas, curvas

cuyo vector tangente es paralelo o derivada covariante nula. Si $\alpha : [a, b] \rightarrow M$ es una curva, tal que $\alpha(a) = p$ y $\alpha(b) = q$, el transporte paralelo es $P_{\alpha(t)} : T_p M \rightarrow T_q M$ un isomorfismo lineal tal que $P_{\alpha(t)}(v) = V(b)$ que es el único transporte paralelo a lo largo de α .

Con la métrica euclidiana la geodésica es el segmento de recta entre dos puntos p y q cualquiera que caracteriza la trayectoria de menor longitud que los une.

Si $\alpha : [a, b] \rightarrow \mathbb{R}^n$ es una curva diferenciable pasando por: $p = \alpha(a)$ y $q = \alpha(b)$, siendo el campo $\frac{d\alpha}{dt}$ asociado físicamente a la velocidad, tenemos la aceleración $\frac{d}{dt}\left(\frac{d\alpha}{dt}\right)$ en cada punto $\alpha(t)$ con la propiedad de α ser geodésica dada por

$$\frac{d}{dt}\left(\frac{d\alpha}{dt}\right) = 0.$$

La extensión de esta noción a variedades exigirá apenas que la componente tangencial de la derivada sea nula.

Geodésicas.

Definición 2.6.5 Una curva parametrizada $\alpha : I \rightarrow M$ es una geodésica si el campo tangente $\frac{d\alpha}{dt}$ verifica:

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

Campos paralelos.

Dado M una variedad diferenciable, una conexión afín ∇ y un campo V a lo largo de una curva diferenciable $\alpha : I \rightarrow M$, V es denominado campo paralelo si $\frac{DV}{dt} = 0$, para todo $t \in I$.

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

Ecuaciones geodésicas.

De la expresión (2.3), un campo paralelo V es determinado por las ecuaciones

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

o, equivalentemente,

$$\frac{dv^k}{dt} + \sum_{i,j=1}^n v_j \frac{d\alpha_i}{dt} \Gamma_{ij}^k = 0, \quad k = 1, \dots, n.$$

Cuando se trata de una geodésica $\alpha(t) = (\alpha_1(t), \dots, \alpha_n(t))$, se tiene $v^i = \frac{d\alpha_i}{dt}$, entonces esta última ecuación se transforma en

$$\frac{d}{dt}\left(\frac{d\alpha_k}{dt}\right) + \sum_{i,j=1}^n \frac{d\alpha_j}{dt} \frac{d\alpha_i}{dt} \Gamma_{ij}^k = 0, \quad k = 1, \dots, n$$

o

$$\frac{d^2\alpha_k}{dt^2} + \sum_{i,j=1}^n \Gamma_{ij}^k \frac{d\alpha_i}{dt} \frac{d\alpha_j}{dt} = 0, \quad k = 1, \dots, n \quad (2.4)$$

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

Conexión afin en variedades riemannianas.

Definición 2.6.6 Sea M una variedad diferenciable con una conexión afin ∇ y una métrica riemanniana \langle, \rangle . Se dice que ∇ es compatible con la métrica \langle, \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:

$$\frac{d}{dt}\langle V, W \rangle = \left\langle \frac{DV}{dt}, W \right\rangle + \left\langle V, \frac{DW}{dt} \right\rangle. \quad (2.5)$$

Proposición 2.6.2 Si la conexión afin ∇ 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 $\alpha(t) = (\alpha_1(t), \dots, \alpha_n(t))$ es una geodésica, $\left\langle \frac{d\alpha}{dt}, \frac{d\alpha}{dt} \right\rangle$ es constante.

Proposición 2.6.3 Sea M una variedad riemanniana. Una conexión afin ∇ es compatible con el \langle, \rangle si, y solamente si:

$$X\langle Y, Z \rangle = \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle, \text{ para todo } X, Y, Z \in \mathcal{H}.$$

Definición 2.6.7 Una conexión afin ∇ en una variedad diferenciable M es llamada simétrica si:

$$\nabla_X Y - \nabla_Y X = [X, Y],$$

donde $[X, Y] = XY - YX$.

Observaciones:

1. En un sistema de coordenadas (U, \mathcal{X}) la simetría de la conexión afín implica

$$\text{que } \nabla_{\frac{\partial}{\partial x_i}} \frac{\partial}{\partial x_j} = \nabla_{\frac{\partial}{\partial x_j}} \frac{\partial}{\partial x_i}.$$

En efecto, para todo $f \in D$,

$$X_i X_j(f) - X_j X_i(f) = \frac{\partial^2 f}{\partial x_i \partial x_j} - \frac{\partial^2 f}{\partial x_j \partial x_i} = 0.$$

2. En consecuencia se tiene que:

$$\nabla_{X_i} X_j - \nabla_{X_j} X_i = \sum_{k=1}^n (\Gamma_{ij}^k - \Gamma_{ji}^k) X_k = 0.$$

Debido a la independencia lineal de $\{X_k\}$ obtenemos:

$$\Gamma_{ij}^k = \Gamma_{ji}^k.$$

La recíproca es inmediata.

El teorema a seguir garantiza la existencia y unicidad de una conexión simétrica y compatible con la métrica en una variedad riemanniana.

Teorema 2.6.1 (*Levi-Civita*). *Dada una variedad riemanniana M , existe una única conexión afín ∇ en M satisfaciendo las condiciones:*

a) ∇ es simétrica.

b) ∇ es compatible con la métrica riemanniana.

(Esta conexión es denominada conexión riemanniana).

Relación entre la métrica riemanniana y los símbolos de Christoffel.

Dado un sistema de coordenadas (U, \mathcal{X}) , las funciones conocidas como símbolos de Christoffel $\Gamma_{i,j}^k : U \rightarrow \mathbb{R}$ definen los coeficientes de conexión $\nabla_{X_i} X_j = \sum_k \Gamma_{ij}^k X_k$. Se muestra que

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

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 su inversa $G^{-1}(x)$ respectivamente.

En efecto, tomemos $\frac{\partial}{\partial x_i} = X_i$, $\frac{\partial}{\partial x_j} = X_j$ y $\frac{\partial}{\partial x_k} = X_k$. Usando el resultado siguiente:

$$\langle Z, \nabla_Y X \rangle = \frac{1}{2} \{ X \langle Y, Z \rangle + Y \langle Z, X \rangle - Z \langle X, Y \rangle - \langle [X, Z], Y \rangle - \langle [Y, Z], X \rangle - \langle [X, Y], Z \rangle \} \quad (2.6)$$

tenemos:

$$\langle X_k, \nabla_{X_j} X_i \rangle = \frac{1}{2} \{ X_i \langle X_j, X_k \rangle + X_j \langle X_k, X_i \rangle - X_k \langle X_i, X_j \rangle \}.$$

Como $\nabla_{X_j} X_i = \nabla_{X_i} X_j = \nabla_{X_i} X_j = \sum_{l=1}^n \Gamma_{ij}^l X_l$ y usando la linealidad del producto interno, se tiene:

$$\sum_{l=1}^n \Gamma_{ij}^l \langle X_k, X_l \rangle = \frac{1}{2} \{ X_i \langle X_j, X_k \rangle + X_j \langle X_k, X_i \rangle - X_k \langle X_i, X_j \rangle \},$$

y así:

$$\sum_{l=1}^n \Gamma_{ij}^l g_{kl} = \frac{1}{2} \left\{ \frac{\partial}{\partial x_i} g_{jk} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right\}.$$

Denotando $b_k = \frac{1}{2} \left\{ \frac{\partial}{\partial x_i} g_{jk} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right\}$, $k = 1, 2, \dots, n$ obtenemos un sistema lineal $Gy = b$ con $y = (\Gamma_{ij}^1, \Gamma_{ij}^2, \dots, \Gamma_{ij}^n)$ y $b = (b_1, b_2, \dots, b_n)$. Como $G(x)$ es invertible (ver definición de métrica riemanniana) entonces $y = G^{-1}b$. Así tenemos:

$$\Gamma_{ij}^m = \frac{1}{2} \sum_{k=1}^n g^{mk} b_k.$$

Finalmente sustituyendo el valor de b_k en la expresión anterior se tiene:

$$\Gamma_{ij}^m = \frac{1}{2} \sum_{k=1}^n \left\{ \frac{\partial}{\partial x_i} g_{jk} + \frac{\partial}{\partial x_j} g_{ki} - \frac{\partial}{\partial x_k} g_{ij} \right\} g^{km}. \quad (2.7)$$

Ejemplo 2.6.1 Sea la variedad riemanniana $M = \mathbb{R}_{++}^n$, con la métrica dada por

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

para funciones $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ diferenciables. La inversa de la matriz $G(x)$ es:

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

1. Obtención de los símbolos de Christoffel.

Recordemos que la relación de la métrica con los símbolos de Christoffel está dado por la ecuación (2.7).

Cuando $k \neq m$ tenemos que $g^{mk} = 0$, así la expresión es reducida a:

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

Consideramos dos casos:

a) Si $i = j$

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

Para $m = i$

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

Para $m \neq i$

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

b) Si $i \neq j$

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

Para $m = i$ entonces, $m \neq j$ y:

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

Para $m = j$ entonces, $m \neq i$ y:

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

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

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

De ambos casos tenemos:

$$\Gamma_{ij}^m = -\frac{1}{h_i(x_i)} \frac{\partial (h_i(x_i))}{\partial x_i} \delta_{im} \delta_{ij} \quad (2.8)$$

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

Como aplicaciones tenemos:

- Si $h_i(x_i) = 1$, entonces, $G(x) = I$. Luego: $\Gamma_{ij}^m = 0, \forall i, j, m = 1, \dots, n$.
- Si $h_i(x_i) = x_i$ entonces, $G(x) = X^{-2}$. Luego: $\Gamma_{ij}^m = -\frac{1}{x_i} \delta_{im} \delta_{ij}$.
- Si $h_i(x_i) = x_i^{\frac{r}{2}}$ entonces, $G(x) = X^{-r}$. Luego: $\Gamma_{ij}^m = -\frac{r}{2} \frac{1}{x_i} \delta_{im} \delta_{ij}$.
- Se $h_i(x_i) = s_i^{-\frac{r}{2}} x_i^{\frac{r}{2}}$, $s_i \in \mathbb{R}_{++}$ entonces, $G(x) = S^r X^{-r}$. Luego $\Gamma_{ij}^m = -\frac{r}{2} \frac{1}{x_i} \delta_{im} \delta_{ij}$.

2. Obtención de la derivada covariante.

Vimos que la relación de la derivada covariante con respecto a los símbolos de Christoffel es dada por la ecuación (2.3). Sustituyendo la expresión (2.8) en (2.3) obtenemos:

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

En particular:

- Si $h_i(x_i) = 1$, $\Gamma_{ij}^k = 0$, y así:

$$\frac{DV}{dt} = \sum_{i=1}^n \frac{dv^i}{dt} X_i,$$

que es la propia derivada usual.

- Si $h_i(x_i) = x_i$, $\Gamma_{ij}^m = -\frac{1}{x_i} \delta_{im} \delta_{ij}$ y

$$\frac{DV}{dt} = \sum_{i=1}^n \left(\frac{dv^i}{dt} - \frac{1}{x_i} v^i \frac{dx_i}{dt} \right) X_i.$$

- Si $h_i(x_i) = x_i^{\frac{r}{2}}$, $\Gamma_{ij}^m = -\frac{r}{2} \frac{1}{x_i} \delta_{im} \delta_{ij}$

$$\frac{DV}{dt} = \sum_{i=1}^n \left(\frac{dv^i}{dt} - \frac{r}{2} \frac{1}{x_i} v^i \frac{dx_i}{dt} \right) X_i.$$

3. Determinación de la ecuación geodésica: Sea $p = (p_1, p_2, \dots, p_n) \in \mathbb{R}_{++}^n$ y $v = (v_1, v_2, \dots, v_n) \in T_p \mathbb{R}_{++}^n = \mathbb{R}^n$ con

$$\alpha : I \rightarrow \mathbb{R}_{++}^n : \alpha(t) = (\alpha_1(t), \alpha_2(t), \dots, \alpha_n(t)),$$

donde $\alpha(0) = p$ y $\frac{d\alpha(0)}{dt} = v$, I algún intervalo abierto de \mathbb{R} . Sustituyendo los símbolos de Christoffel (2.8) en la ecuación (2.4) obtenemos:

$$\frac{d^2 \alpha_i}{dt^2} - \frac{1}{h_i(\alpha_i)} \frac{\partial(h_i(\alpha_i))}{\partial \alpha_i} \left(\frac{d\alpha_i}{dt} \right)^2 = 0, \forall i = 1, \dots, n \quad (2.9)$$

$$\begin{aligned}\alpha_i(0) &= p_i, & i &= 1, \dots, n \\ \alpha'_i(0) &= v_i, & i &= 1, \dots, n.\end{aligned}$$

La ecuación diferencial (2.9) es equivalente a resolver:

$$\frac{d\alpha_i}{dt} = h_i(\alpha_i)a_i,$$

para alguna constante a_i , que también es equivalente a resolver la integral:

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

para algunas constantes a_i y b_i en \mathbb{R} .

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

$$\int \left(\frac{1}{h_i(\alpha_i)} \right) d\alpha_i = a_i t + b_i \quad i = 1, \dots, n \quad (2.10)$$

donde a_i y b_i son constantes reales tales que:

$$\begin{aligned}\alpha_i(0) &= p_i, & i &= 1, \dots, n. \\ \alpha'_i(0) &= v_i, & i &= 1, \dots, n.\end{aligned}$$

En particular:

- Si $h_i(\alpha_i) = 1$ tenemos que $G(p) = I$ y considerando las condiciones iniciales de (2.10) encontramos la expresión de la curva geodésica:

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

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

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

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

- Si $h(\alpha_i) = \alpha_i$ entonces, $G(x) = X^{-2}$ considerando las condiciones iniciales de (2.10), las curvas geodésicas son funciones exponenciales:

$$\alpha(t) = \left(p_1 \exp\left(\frac{v_1}{p_1} t\right), p_2 \exp\left(\frac{v_2}{p_2} t\right), \dots, p_n \exp\left(\frac{v_n}{p_n} t\right) \right).$$

Vemos que dados cualquier $p \in \mathbb{R}_{++}^n$ y $v \in \mathbb{R}^n$, la geodésica $\alpha(t)$ está definida para todo $t \in \mathbb{R}$.

Ejemplo 2.6.2 Consideremos la variedad riemanniana $C_0^n = (0, 1)^n$ con la métrica dada por:

$$G(p) = \text{diag} \left(\frac{1}{(h_1(\alpha_1))^2}, \frac{1}{(h_2(\alpha_2))^2}, \dots, \frac{1}{(h_n(\alpha_n))^2} \right),$$

para funciones $h_i : (0, 1) \rightarrow (0, 1)$ diferenciables. Así, la única geodésica $\alpha(t)$ de C_0^n , con métrica $G(p)$, pasando por el punto $\alpha(0) = p$, en la dirección $\alpha'(0) = v \in T_p C_0^n = \mathbb{R}^n$, es obtenida resolviendo la siguiente ecuación:

$$\int \left(\frac{1}{h_i(\alpha_i)} \right) d\alpha_i = a_i t + b_i \quad i = 1, \dots, n,$$

donde a_i y b_i son constantes reales tales que:

$$\alpha_i(0) = p_i, \quad i = 1, \dots, n.$$

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

- Si $h_i(\alpha_i) = \text{sen}^2(\pi\alpha_i)$ entonces $G(p) = \text{csc}^4(\pi p)$, con las geodésicas en C_0^n ,

$$\alpha(t) = (\alpha_1(t), \alpha_2(t), \dots, \alpha_n(t)),$$

considerando las condiciones iniciales de (2.10), son:

$$\alpha_i(t) = \frac{1}{\pi} \text{arccot} \left(-\pi \text{csc}^2(\pi p_i) v_i t + \cot(\pi p_i) \right) \text{ para todo, } i = 1, 2, \dots, n.$$

Observamos que dados cualquier $p \in C_0^n$ y $v \in \mathbb{R}^n$, la geodésica $\alpha(t)$ está definida para todo $t \in \mathbb{R}$.

- Si $h_i(\alpha_i) = \alpha_i(1 - \alpha_i)$ se tiene que $G(p) = P^{-2}(I - P)^{-2}$, las geodésicas en C_0^n ,

$$\alpha(t) = (\alpha_1(t), \alpha_2(t), \dots, \alpha_n(t)),$$

considerando las condiciones iniciales de (2.10), son:

$$\alpha_i(t) = \frac{1}{2} \left\{ 1 + \text{tgh} \left(\frac{1}{2} \frac{v_i}{p_i(1 - p_i)} t + \text{arccoth}(2p_i - 1) \right) \right\} \text{ para todo, } i = 1, 2, \dots, n.$$

donde $\text{tanh}(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ es la función tangente hiperbólica.

Observamos que dados cualesquiera $p \in C_0^n$ y $v \in \mathbb{R}^n$, la geodésica $\alpha(t)$ está definida para todo $t \in \mathbb{R}$.

2.7 Curvatura de una variedad riemanniana

En esta sección, presentamos la definición de curvatura de una variedad riemanniana que, intuitivamente, mide cuanto ella se aleja de ser euclidiana. Del punto de vista de las aplicaciones esta sección muestra esencialmente que las variedades \mathbb{R}_{++}^n y C_0^n con la métrica dada por $G(x) = \text{diag} \left(\frac{1}{(h_1(x_1))^2}, \frac{1}{(h_2(x_2))^2}, \dots, \frac{1}{(h_n(x_n))^2} \right)$ para cualquier función diferenciable $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$ y $h_i : (0, 1)^n \rightarrow \mathbb{R}_{++}$ respectivamente, tiene curvatura cero.

Definición 2.7.1 (*Aplicación curvatura*). Sea $A(\mathcal{H}, \mathcal{H})$ el conjunto de aplicaciones de \mathcal{H} en \mathcal{H} y ∇ la conexión afín en una variedad riemanniana M , dada por el teorema de Levi-Civita.

Una curvatura K de una variedad riemanniana M es una correspondencia

$$K : \mathcal{H} \times \mathcal{H} \rightarrow A(\mathcal{H}, \mathcal{H})$$

definida por:

$$K(X, Y)Z = \nabla_Y \nabla_X Z - \nabla_X \nabla_Y Z + \nabla_{[X, Y]} Z.$$

Observación 2.7.1 Si la variedad $M = \mathbb{R}^n$, entonces $K(X, Y)Z = 0$, para todo $X, Y, Z \in \mathcal{H}$. En efecto, basta indicar $Z = (z_1, z_2, \dots, z_n)$ las componentes del campo Z en las coordenadas naturales de \mathbb{R}^n y la conexión definida por:

$$\nabla_X Z = (X z_1, X z_2, \dots, X z_n).$$

Observación 2.7.2 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:

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

Observación 2.7.3 La curvatura K es antisimétrica. En efecto,

$$K(X, Y)Z + K(Y, X)Z = \nabla_{[X, Y]} Z + \nabla_{[Y, X]} Z, \text{ para todo } Z \in \mathcal{H}.$$

Como $[X, Y] = -[Y, X]$, entonces:

$$K(X, Y)Z + K(Y, X)Z = 0, \text{ para todo } Z \in \mathcal{H},$$

y así,

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

Proposición 2.7.1 *La curvatura K de una variedad riemanniana es trilineal, en el siguiente sentido:*

a. 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 $f, g \in D(M)$ y $X_1, X_2, Y_1, Y_2 \in \mathcal{H}$.

b. Para todo par $X, Y \in \mathcal{H}$, el operador curvatura $K(X, Y) : \mathcal{H} \rightarrow \mathcal{H}$ 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 $f \in D(M)$ y $Z, W \in \mathcal{H}$.

Proposición 2.7.2 *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 las componentes 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 2.7.4 *Si en las coordenadas (U, \mathcal{X}) escribimos: $X = \sum_{i=1}^n u^i X_i$, $Y = \sum_{j=1}^n v^j X_j$, $Z = \sum_{k=1}^n w^k X_k$, por la linealidad de K tenemos:*

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

Ejemplo 2.7.1 *Sea $M = \mathbb{R}_{++}^n$ ó $M = C_0^n$ con estructura de variedad riemanniana representada por la matriz $G(x) = \text{diag}(\frac{1}{(h_1(x_1))^2}, \frac{1}{(h_2(x_2))^2}, \dots, \frac{1}{(h_n(x_n))^2})$. Ya vimos que sus símbolos de Christoffel son:*

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

Si en las coordenadas (U, \mathcal{X}) escribimos: $X = \sum_{i=1}^n u^i X_i$, $Y = \sum_{j=1}^n v^j X_j$, $Z = \sum_{k=1}^n w^k X_k$, de la tri-linealidad de K tenemos:

$$K(X, Y)Z = \sum_{i,j,k=1}^n u^i v^j w^k K(X_i, X_j)X_k.$$

Por definición de curvatura:

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

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 $i \neq j$, entonces

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

Sustituyendo los símbolos de Christoffel tenemos:

$$\nabla_{X_i}X_k = \sum_{j=1}^n \left(-\frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \delta_{ij} \delta_{ik} \right) X_j = -\frac{1}{h_i(x_i)} \frac{\partial(h_i(x_i))}{\partial x_i} \delta_{ik} X_i \quad (2.11)$$

luego tomando ∇_{X_j} se tiene:

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

por definición de conexión afín $\nabla_X(fY) = f\nabla_X Y + X(f)Y$ donde $X(f) = \sum_{i=1}^n a_i(\cdot) \frac{\partial f}{\partial x_i}$, entonces tenemos:

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

Usando (2.11) y dado que $i \neq j$, el primero y segundo término de la suma anterior, es igual a cero. Por tanto:

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

Análogamente,

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

De ambos resultados se tiene:

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

Así $K(X, Y)Z = 0$. Luego las variedades riemannianas \mathbb{R}_{++}^n y C_0^n con métrica $G(x)$ tienen curvatura cero. En particular, con las métricas I, X^{-r} , para \mathbb{R}_{++}^n y $\text{cosec}^4(\pi x)$, $X^{-r}(I - X)^{-r}$ para C_0^n , son variedades de curvatura cero.

Curvatura Seccional.

Intimamente relacionado con el operador curvatura K está la curvatura seccional (o riemanniana) que definiremos a seguir.

Sea M una variedad riemanniana y σ un subespacio bidimensional de T_pM . Definimos la forma cuadrática como $Q : \sigma \rightarrow \mathbb{R}$ tal que:

$$Q(x, y) = \langle x, x \rangle \langle y, y \rangle - \langle x, y \rangle^2.$$

Geométricamente $\sqrt{Q(x, y)}$ representa el área del paralelogramo definido por x e y .

Proposición 2.7.3 Sea $\sigma \subset T_pM$ un subespacio bidimensional y $x, y \in \sigma$, dos vectores linealmente independientes. Entonces,

$$K(x, y) = \frac{\langle K(x, y)x, y \rangle}{Q(x, y)},$$

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

Definición 2.7.2 (Curvatura Seccional). Dado un punto $p \in M$ y $\sigma \subset T_pM$. El número $K(x, y) = K(\sigma)$, donde $\{x, y\}$ es una base de σ , 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.

2.8 Gradiente y Hessiana en una variedad riemanniana

Sea M una variedad riemanniana y $f : M \rightarrow \mathbb{R}$ una función diferenciable. Dado $p \in M$ sabemos que la diferencial de f en el punto p es un funcional lineal definido en T_pM , entonces por el teorema de representación de Riesz existe un único elemento denotado por $\nabla_M f(p) \in T_pM$ tal que para todo $v \in T_pM$ se tiene

$$df_p(v) = \langle \nabla_M f(p), v \rangle \quad (2.12)$$

y

$$\|\nabla_M f(p)\| = \|df_p\|,$$

esto es, la aplicación diferencial se puede caracterizar por la aplicación de producto interno. Así podemos definir un campo vectorial $grad f : M \rightarrow TM$, como

$$grad f(p) = \nabla_M f(p).$$

La expresión (2.12) puede ser escrita como:

$$df_p(X(p)) = \langle grad f(p), X(p) \rangle, \text{ para todo } X \in \mathcal{H},$$

y así también podemos definir una aplicación $df : \mathcal{H} \rightarrow M^* = \mathcal{L}(M, \mathbb{R})$, donde $\mathcal{L}(M, \mathbb{R})$ es el conjunto de funciones en M en \mathbb{R} , tal que:

$$df(X) = \langle grad f, X \rangle.$$

Además, $df_p(X(p)) = \frac{d}{dt}(f \circ \gamma)|_{t=0}$ para alguna curva $\gamma : I \rightarrow M$ con $\gamma(0) = p$ y $\gamma'(0) = X(p)$, luego tenemos que $df_p(X(p)) = \sum_{i=1}^n \gamma'_i(0) \frac{\partial f}{\partial x_i}(p) = X(f)(p)$, por tanto

$$df(X) = \langle grad f, X \rangle = X(f).$$

Así llegamos a la siguiente definición.

Definición 2.8.1 *El gradiente de una función diferenciable $f : M \rightarrow \mathbb{R}$ es un campo vectorial $grad f : M \rightarrow TM$ métricamente equivalente a la diferencial, esto es,*

$$df_p(X(p)) = \langle grad f(p), X(p) \rangle = X(p)f, \text{ para todo } X \in \mathcal{H}.$$

Observación 2.8.1 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:

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

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

$$df_q(v) = f'(q)^T v = f'(q)^T (G(q)^{-1})^T G(q) v = (G(q)^{-1} f'(q))^T G(q) v = \langle G(q)^{-1} f'(q), v \rangle_q.$$

Ejemplo 2.8.1 Sea la variedad riemanniana \mathbb{R}^n con la métrica euclidiana $G = I$, entonces $\text{grad } f(x) = f'(x)$ (el gradiente usual).

Ejemplo 2.8.2 Sea la variedad riemanniana \mathbb{R}_{++}^n con la expresión de la métrica

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

para funciones $h_i : \mathbb{R}_{++} \rightarrow \mathbb{R}_{++}$

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

En particular:

1. Si $h_i(x_i) = x_i$ entonces:

$$\text{grad } f(x) f(x) = X^2 f'(x),$$

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

2. Si $h_i(x_i) = x_i^{\frac{r}{2}}$, $r \neq 2$ entonces:

$$\text{grad } f(x) f(x) = X^r f'(x).$$

Ejemplo 2.8.3 Sea la variedad riemanniana $C_0^n = (0, 1)^n$ con la expresión de la métrica dada por $G(x) = \csc^4(\pi x) = \text{diag}(\csc^4(x_1), \dots, \csc^4(x_n))$, entonces

$$\text{grad } f(x) f(x) = \text{sen}^4(\pi x) f'(x),$$

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

Definición 2.8.2 Sea M una variedad riemanniana y $p \in M$. Decimos que p es punto crítico si $\text{grad } f(p) = 0$.

Definición 2.8.3 (Hessiana de una función). Sea $f : M \rightarrow \mathbb{R}$ 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 gradiente, esto es,

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

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

$$H_p^f(v) = \frac{D}{dt}(\text{grad } f)(p) = \nabla_v \text{grad } f(p).$$

A partir del concepto de Hessiana podemos definir las aplicaciones $H_p^f : T_p M \rightarrow T_p M$ y $H^f : M \rightarrow \mathcal{L}(TM, TM)$ donde $\mathcal{L}(TM, TM)$ es el conjunto de aplicaciones lineales de TM en TM y $H^f(p) = H_p^f \in \mathcal{L}(T_p M, T_p M)$.

Proposición 2.8.1 Para cada $p \in M$, el operador $H_p^f : T_p M \rightarrow T_p M$ es lineal y autoadjunto, esto es, $\langle H_p^f(v), w \rangle_p = \langle v, H_p^f(w) \rangle_p$.

De la Proposición anterior, para cada $p \in M$ podemos introducir una forma cuadrática $q_p^f : T_p M \times T_p M \rightarrow \mathbb{R}$ definida por:

$$q_p^f(v, w) = \langle H_p^f v, w \rangle_p.$$

Más generalmente, podemos definir la aplicación $q^f : \mathcal{H} \times \mathcal{H} \rightarrow \mathcal{L}(M, \mathbb{R})$ dada por:

$$q^f(X, Y) = \langle \nabla_X \text{grad } f, Y \rangle. \quad (2.13)$$

La función definida en (2.13) tiene la desventaja de depender del conocimiento de la métrica y de la conexión, cuando sabemos que la métrica determina una conexión afín (Teorema de Levi Civita), por tanto la proposición siguiente es importante para poder obtener una caracterización adecuada.

Proposición 2.8.2 Para todo $X, Y \in \mathcal{H}$

$$q^f(X, Y) = (XY - \nabla_X Y)f = (YX - \nabla_Y X)f.$$

Observación 2.8.2 En un sistema de coordenadas (\mathcal{X}, U) en terminos de la base $\{X_k\}$ tenemos:

$$q^f(X_i, X_j) = \left(X_i X_j - \sum_{m=1}^n \Gamma_{ij}^m X_m \right) f,$$

esto es:

$$q^f(X_i, X_j) = \langle H_p^f X_i, X_j \rangle = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} - \sum_{m=1}^n \Gamma_{ij}^m \frac{\partial f}{\partial x_m} \right). \quad (2.14)$$

Ejemplo 2.8.4 Sea la variedad riemanniana \mathbb{R}^n con métrica $G(x) = I$; como vimos anteriormente, los símbolos de Christoffel son $\Gamma_{ij}^m = 0$, para todo $i, j, m = 1, \dots, n$, entonces la matriz Hessiana es la Hessiana usual $H_p^f(p) = f''(p)$.

Ejemplo 2.8.5 Sea la variedad riemanniana \mathbb{R}_{++}^n con la métrica $g_{ij} = \frac{\delta_{ij}}{h_i(x_i)h_j(x_j)}$. Sabemos que los símbolos de Christoffel son:

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

entonces:

$$q^f(X_i X_j) = X_i X_j + \sum_{m=1}^n \frac{1}{h_i(x_i)} \frac{\partial h_i(x_i)}{\partial x_i} \delta_{im} \delta_{ij} X_m.$$

Si $m \neq j$ entonces $\delta_{im} \delta_{ij} = 0$, luego se tiene:

$$q^f(X_i, X_j) = \left(\frac{\partial^2}{\partial x_i \partial x_j} + \frac{1}{h_i(x_i)} \frac{\partial h_i(x_i)}{\partial x_i} \delta_{ij} \frac{\partial}{\partial x_i} \right) f.$$

Así, $H_x^f = (q^f(X_i)(X_j))$ es la matriz que representa la Hessiana de la función f . Aún podemos dar una representación matricial

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

donde:

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

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

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

En particular:

1. Si $h_i(x_i) = 1$ entonces $H_x^f = f''(x)$ la matriz Hessiana usual.
2. Si $h_i(x_i) = x_i$ entonces $H_x^f = f''(x) + X^{-1}\mathcal{F}'(x)$.
3. Si $h_i(x_i) = x_i^{\frac{r}{2}}$, $r \neq 2$, entonces: $H_x^f = f''(x) + \frac{r}{2}X^{-1}\mathcal{F}'(x)$.

Corolario 2.8.1 Si $p \in M$ es un punto crítico de f y $X, Y \in \mathcal{H}$, entonces:

$$H_p^f(X(p), Y(p)) = X(p)Y(p)f.$$

Demostración. $H_p^f(X(p), Y(p)) = X(p)(Y(p)f) - \langle \nabla_{X(p)}Y(p), \text{grad } f(p) \rangle$, y como $\text{grad } f(p) = 0$, se sigue el Corolario. ■

De este corolario, se deduce que si $p \in M$ es un punto crítico de f entonces la matriz Hessiana de f , calculada en este punto, coincide con la matriz Hessiana usual.

2.9 Variedades completas

Todos los algoritmos desarrollados en Optimización, en la perspectiva de la geometría riemanniana, necesitan de la hipótesis de variedad completa, que en términos simples significa que la geodésica, contenida en ella, está definida para todos los valores de $t \in \mathbb{R}$. El Teorema de Hopf y Rinow de gran importancia y utilidad en las aplicaciones el cual enunciaremos posteriormente, dice: Dados dos puntos cualquiera de la variedad completa existe siempre una geodésica que minimiza la longitud de arco entre todas las curvas regulares por partes que unen tales puntos.

Una pregunta natural sería, si se podría desarrollar métodos geodésicos donde la hipótesis de variedad completa sea más suave, por ejemplo, introducir una medida en la variedad a partir del producto interno del espacio tangente donde la geodésica esté definida en casi todos los puntos y los puntos donde la geodésica no sea definida, pertenezca a un conjunto de medida nula.

Restringiéndonos a las variedades completas, desarrollamos esta teoría de manera resumida.

Definición 2.9.1 Una variedad riemanniana M es llamada (geodésicamente) completa si para todo $p \in M$, las geodésicas que parten de p están definidas para todos los valores del parámetro $t \in \mathbb{R}$.

Ejemplo 2.9.1 \mathbb{R}^n con la métrica euclidiana $G(x) = I$, es (geodesicamente) completa, pues dado un punto cualquiera $x \in M$ y una dirección arbitraria $v \in T_x M$, vemos que la i -ésima componente de la geodésica que cumple las condiciones iniciales $\alpha_i(0) = x_i$ y $\alpha'_i(0) = v_i$, para todo $i = 1, \dots, n$, es dado por: $\alpha_i(t) = x_i + tv_i$, para todo $i = 1, 2, \dots, n$, lo que está definida para todo $t \in \mathbb{R}$.

Ejemplo 2.9.2 \mathbb{R}_{++}^n con la métrica $G(x) = X^{-2}$ es (geodésicamente) completa, pues dado un punto cualquiera $x \in M$ y una dirección cualquiera $v \in T_x M$, vemos que la i -ésima componente de la geodésica cumpliendo las condiciones iniciales $\alpha_i(0) = x_i$ y $\alpha'_i(0) = v_i$, para todo $i = 1, \dots, n$ es : $\alpha_i(t) = x_i \exp(\frac{v_i}{x_i} t)$ la cual está definido para todo $t \in \mathbb{R}$.

Ejemplo 2.9.3 $C_0^n = (0, 1)^n$ con la métrica dada por

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

es (geodesicamente) completa pues dado un punto arbitrario $x \in M$ y una dirección cualquiera $v \in T_x M$, vemos que la i -ésima componente de la geodésica cumpliendo las condiciones iniciales $\alpha_i(0) = x_i$ y $\alpha'_i(0) = v_i$, para todo $i = 1, \dots, n$ es:

$$\alpha_i(t) = \frac{1}{\pi} \text{arccot} \left(-\pi \text{csc}^2(\pi x_i) v_i t + \cot(\pi x_i) \right),$$

para todo $i = 1, 2, \dots, n$.

En adelante consideramos que la variedad riemanniana estudiada tiene la propiedad de conexidad, esto es, para cualquier par de puntos p, q de M existe una curva diferenciable contenida en M , $\gamma : [a, b] \rightarrow M$, tal que $\gamma(a) = p$ y $\gamma(b) = q$.

Definición 2.9.2 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

$$d(p, q) = \inf_{\gamma} \int_a^b \|\gamma'(t)\| dt \quad (2.15)$$

donde $\gamma : [a, b] \rightarrow M$ es una curva diferenciable tal que $\gamma(a) = p$ y $\gamma(b) = q$.

Proposición 2.9.1 Con la distancia geodésica (2.15) M es un espacio métrico.

Teorema 2.9.1 (Hopf-Rinow) Sea M una variedad riemanniana y sea $p \in M$. Las siguientes afirmaciones son equivalentes:

- a) Limitados y cerrados son compactos.
- b) M es completo como espacios métrico.
- c) M es geodesicamente completa.
- d) Para todo $q \in M$ existe una geodésica uniendo p y q con:

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

esto es, el mínimo de (2.15) es obtenida por una geodésica.

Ejemplo 2.9.4 Sea la variedad riemanniana \mathbb{R}_{++}^n con expresión de la métrica $G(x) = X^{-2}$. Dados p y q en \mathbb{R}_{++}^n , existe una única geodésica uniendo p a q . En efecto, sea $\gamma_i(t) = p_i \exp(\frac{v_i t}{p_i})$ y $\alpha_i(t) = p_i \exp(\frac{w_i t}{p_i})$ las i -ésimas componentes que satisfacen las condiciones iniciales:

$$\gamma_i(0) = \alpha_i(0) = p_i,$$

$$\gamma_i(t_0) = \alpha_i(t_0) = q_i.$$

Se puede verificar que $v_i = w_i$ para todo $i = 1, \dots, n$. En efecto, tomando el valor $t = t_0$ tenemos que para todo $i = 1, \dots, n$: $q_i = p_i \exp(\frac{v_i t_0}{p_i}) = p_i \exp(\frac{w_i t_0}{p_i})$ dividiendo por p_i , tomando logaritmo y multiplicando por p_i/t_0 tenemos $v_i = w_i$ y así $\gamma_i(t) = \alpha_i(t)$ para todo $t \in \mathbb{R}$. Además:

a). Debido a que $q_i = p_i \exp(\frac{v_i}{p_i})$ entonces $v_i = p_i \ln(\frac{q_i}{p_i})$.

b). $\frac{\gamma_i'(t)}{\gamma_i(t)} = \frac{v_i}{p_i}$ entonces $\left(\frac{\gamma_i'(t)}{\gamma_i(t)}\right) = \frac{1}{t_0} \ln^2\left(\frac{q_i}{p_i}\right)$ así:

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

Ejemplo 2.9.5 Sea la variedad riemanniana $C_0^n = (0, 1)$ con la expresión de la métrica riemanniana $G(x) = \csc^4(\pi x)$. Dados p y q en C_0^n , existe una única geodésica que une p y q . Además tenemos:

a). $q_i = \gamma_i(t_0) = \frac{1}{\pi} \operatorname{arccot}(-\pi \csc^2(\pi p_i) v_i t_0 + \operatorname{ctg}(\pi p_i))$, para todo $i = 1, \dots, n$, entonces

$$v_i = -\frac{\cot(\pi q_i) - \cot(\pi p_i)}{\pi t_0 \csc^2(\pi p_i)}.$$

b). $\frac{\gamma_i'(t)}{\operatorname{sen}^2(\pi \gamma_i(t))} = a_i = \csc^2(\pi p_i) v_i$. Substituyendo el valor de v_i y elevando al cuadrado tenemos:

$$\left(\frac{\gamma_i'(t)}{\operatorname{sen}^2(\pi \gamma_i(t))} \right)^2 = a_i^2 = \frac{1}{\pi^2 t_0^2} \{ \cot(\pi q_i) - \cot(\pi p_i) \}^2$$

usando estos hechos tenemos:

$$d(p, q) = \int_0^{t_0} \|\gamma'(t)\| dt = \frac{1}{\pi} \left\{ \sum_{i=1}^n [\cot(\pi q_i) - \cot(\pi p_i)]^2 \right\}^{\frac{1}{2}}.$$

Teorema 2.9.2 *Ley de cosenos.* Sea M una variedad riemanniana completa con curvatura seccional no negativa, en un triángulo geodésico normalizado tal que $\gamma_1, \gamma_2, \gamma_3$ segmento de geodésicas minimizantes. Vale la desigualdad

$$c^2 \leq a^2 + b^2 - 2ab \cos \alpha \tag{2.16}$$

donde $\alpha = \arg(\gamma_1'(0) \cdot -\gamma_3'(l_3))$, $a = L(\gamma_1)$, $b = L(\gamma_3)$, $c = L(\gamma_2)$, L , longitud de geodésica

Capítulo 3

El Problema de Optimización y sus Condiciones de Optimalidad

Diversos fenómenos naturales son estudiados a través de modelos matemáticos, en especial por modelos presentados por un problema de optimización el cual mostraremos más adelante. En estos modelos, es necesario garantizar inicialmente las condiciones para la existencia y caracterización de puntos óptimos, para luego desarrollar un algoritmo adecuado que resuelva algunos modelos matemáticos de optimización, en tal sentido, definimos el conjunto sobre el cual estamos trabajando, es decir conoceremos lo que es una variedad convexa y en ella estudiamos una clase particular de funciones llamadas convexas y cuasi-convexas.

Para el desarrollo de este Capítulo, iniciamos con algunas definiciones elementales, que serán de gran utilidad al resolver un problema de optimización.

Definición 3.0.3 (*Mínimo: global, local, estricto*). Sea M una variedad riemanniana completa y $f : M \rightarrow \mathbb{R}$ una función.

1. $\bar{x} \in M$ es un mínimo global de f si, $f(\bar{x}) \leq f(x)$, para todo $x \in M$.
2. $\bar{x} \in M$ es un mínimo local de f si, existe $\delta > 0$ tal que:

$$f(\bar{x}) \leq f(x), \text{ para todo } x \in B(\bar{x}, \delta),$$

donde $B(\bar{x}, \delta) = \{x \in M, d(\bar{x}, x) < \delta\}$.

3. $\bar{x} \in M$ es mínimo local estricto si, existe $\delta > 0$ tal que $f(\bar{x}) < f(x)$, para todo $x \neq \bar{x}$, $x \in B(\bar{x}, \delta)$.

El problema de interés será resolver el siguiente modelo:

$$\begin{aligned} \min f(x) \\ x \in M \end{aligned} \tag{3.1}$$

que significa encontrar los mínimos globales de una función f sobre M , y es denominado “Problema de Minimización”, sujeta generalmente a algunas restricciones sobre su dominio.

El modelo (3.1) resuelve también el “Problema de Maximización”.

$$\begin{aligned} \max f(x) \\ x \in M \end{aligned}$$

para ello basta definir $f(x) = -g(x)$.

3.1 Existencia de puntos de mínimo global

Definición 3.1.1 Una función $f : M \rightarrow \mathbb{R}$ es denominada semicontinua inferior en $\bar{x} \in M$, si para toda sucesión $\{x^k\}$ de M convergente a \bar{x} se tiene que:

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

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

El siguiente Teorema garantiza la existencia de un punto de mínimo global para el problema (3.1).

Teorema 3.1.1 (Weierstrass) Considere el problema (3.1), si $f : M \rightarrow \mathbb{R}$ es semicontinua inferior y M es compacto, entonces existe un punto de mínimo global de f .

Demostración. Mostraremos inicialmente que f es limitada inferiormente, esto es, existe $\alpha \in \mathbb{R}$ tal que:

$$\alpha \leq f(x), \text{ para todo } x \in M.$$

Por contradicción, supongamos que f no es limitada inferiormente, entonces existe una sucesión $\{x^k\} \subset M$ tal que:

$$\lim_{k \rightarrow +\infty} f(x^k) = -\infty. \quad (3.2)$$

Dado que M es compacto, entonces existe una subsucesión $\{x^{k_j}\} \subset \{x^k\}$ tal que:

$$\lim_{j \rightarrow +\infty} x^{k_j} = \hat{x},$$

por la semicontinuidad inferior de f tenemos:

$$\liminf_{j \rightarrow +\infty} f(x^{k_j}) \geq f(\hat{x}),$$

lo que contradice a (3.2), por lo tanto f es limitada inferiormente en M . De aquí existe $f^* \in \mathbb{R}$ tal que $f^* = \inf\{f(x) : x \in M\}$. Por propiedad de ínfimo, existe una sucesión $\{x^k\} \subset M$ tal que:

$$\lim_{k \rightarrow \infty} f(x^k) = f^*.$$

Por la compacidad de M , existe \bar{x} y $\{x^{k_j}\} \subset \{x^k\}$ tal que $\lim_{j \rightarrow \infty} x^{k_j} = \bar{x} \in M$. Nuevamente, por la semicontinuidad inferior de f

$$\liminf_{j \rightarrow \infty} f(x^{k_j}) \geq f(\bar{x}).$$

Como $\{f(x^k)\}$ converge a f^* , la subsucesión $\{f(x^{k_j})\}$ converge a f^* obteniendo que

$$f^* \geq f(\bar{x}),$$

así, \bar{x} es un punto de mínimo global de f en M . ■

3.2 Caracterización de puntos de mínimo local

Presentamos en esta Sección las condiciones necesarias y suficientes de optimalidad local para el problema (3.1).

Teorema 3.2.1 (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. Tomemos $v \in T_{x^*}M$ y una curva geodésica $\gamma : \mathbb{R} \rightarrow M$ con condiciones $\gamma(0) = x^*$ 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 para f , entonces existe $\delta > 0$ tal que

$$h(0) = f(x^*) \leq f(\gamma(t)) = h(t),$$

para todo $t \in (-\delta, \delta)$ lo que implica que en $t = 0$ tenemos un punto de mínimo local de h . Por la condición necesaria de primer orden en \mathbb{R} se tiene

$$h'(0) = \langle \text{grad } f(x^*), v \rangle = 0.$$

Tomando en particular $v = \text{grad } f(x^*)$ tenemos que $\text{grad } f(x^*) = 0$. ■

Teorema 3.2.2 (Condición necesaria de segundo orden). Sea $f : M \rightarrow \mathbb{R}$ de clase C^2 . Si x^* es punto de mínimo local, entonces $\langle v, H_{x^*}^f v \rangle \geq 0$, $\forall v \in T_{x^*}M$.

Demostración. Sea $v \in T_{x^*}M$, y $\gamma : \mathbb{R} \rightarrow M$ una geodésica con $\gamma(0) = x^*$, $\gamma'(0) = v$. Definimos $h : \mathbb{R} \rightarrow \mathbb{R}$ tal que $h(t) = f(\gamma(t))$. Del Teorema 3.2.1, en $t = 0$ tenemos un punto de mínimo local de h , entonces por la condición necesaria de segundo orden: $h'(0) = 0$, luego $h''(0) \geq 0$.

Veamos:

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

Teorema 3.2.3 (Condición suficiente de segundo orden). Sea $f : M \rightarrow \mathbb{R}$ de clase C^2 . Si $x^* \in M$ que satisface:

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

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

Demostración. Por contradicción. Supongamos que x^* no es punto de mínimo local estricto, entonces existe una subsucesión $\{x^k\} \in B(x^*, \frac{1}{k}) \setminus \{x^*\}$ tal que:

$$f(x^*) \geq f(x^k). \quad (3.3)$$

Sea la geodésica minimal $\gamma_k : [0, 1] \rightarrow \mathbb{R}$ tal que $\gamma_k(0) = x^*$, $\gamma_k(1) = x^k$, $\gamma_k'(0) = v_k$ y $d(x^*, x^k) = \|\exp_{x^*} x^k\|$. Definimos $h : \mathbb{R} \rightarrow \mathbb{R}$ tal que $h(t) = (f \circ \gamma_k)(t)$ y por el desarrollo de Taylor de segundo orden de h en 0 :

$$h(t) = h(0) + th'(0) + \frac{1}{2}t^2h''(0) + \theta(|t|^2), \quad \text{donde, } \lim_{t \rightarrow 0} \frac{\theta(|t|^2)}{|t|^2} = 0,$$

esto es,

$$f(\gamma_k(t)) = f(x^*) + \frac{t^2}{2} \left\langle \frac{D}{dt} \text{grad } f(x^*) v_k, v_k \right\rangle + \theta(|t|^2).$$

Evalutando en $t = 1$

$$f(x^k) = f(x^*) + \frac{1}{2} \langle v_k, H_{x^*}^f v_k \rangle + \theta(d^2(x^*, x^k)) \quad (3.4)$$

donde: $\lim_{d \rightarrow 0} \frac{\theta(d^2(x^*, x^k))}{d^2(x^*, x^k)} = 0$.

Definamos $z^k = \frac{v_k}{\|v_k\|}$, la sucesión $\{z^n\}$ es limitada, entonces existe una subsucesión $\{z^{k_j}\} \subset \{z^k\}$ tal que $\{z^{k_j}\} \rightarrow \bar{z}$. Substituyendo en (3.4) k por k_j , tenemos:

$$f(x^{k_j}) = f(x^*) + \frac{1}{2} \langle v_{k_j}, H_{x^*}^f v_{k_j} \rangle + \theta(d^2(x^*, x^{k_j})) \quad (3.5)$$

donde: $\lim_{d \rightarrow 0} \frac{\theta(d^2(x^*, x^{k_j}))}{d^2(x^*, x^{k_j})} = 0$.

De la relación (3.3) y tomando límite en (3.5) cuando $j \rightarrow \infty$, obtenemos:

$$0 \geq \langle \bar{z}, H_{x^*}^f \bar{z} \rangle,$$

lo que contradice la hipótesis b) del Teorema 3.2.3. Por tanto, x^* es un punto de mínimo local estricto. ■

3.3 Elementos del análisis convexo

La teoría del análisis convexo en variedades riemannianas fueron estudiadas por RAPCSÁK [20] y UDRISTE [22]. Rapsák considera una variedad diferenciable con

métrica inducida de \mathbb{R}^n , obteniendo caracterizaciones de primer y segundo orden. UDRISTE consideró el estudio sobre una variedad riemanniana abstracta generalizando (independientemente) la teoría de convexidad.

A partir de estos trabajos el estudio del análisis convexo con aplicación a la teoría de Optimización se ha profundizado, véase por ejemplo da CRUZ NETO et. al.[5] y [6], FERREIRA y OLIVEIRA[9] quienes consideran en sus estudios, una variedad riemanniana completa con curvatura seccional no negativa, bajo esta misma perspectiva, desarrollamos básicamente el análisis convexo sobre una variedad riemanniana.

En esta Sección damos algunas nociones de los elementos del análisis convexo y con esta misma perspectiva definimos funciones convexas y cuasi-convexas en una variedad riemanniana.

3.3.1 Convexidad en una variedad riemanniana

Existen diversos puntos de vista en la geometría riemanniana para generalizar el concepto de convexidad de \mathbb{R}^n , los más importantes son los que presentamos en las siguientes definiciones.

Definición 3.3.1 *Sea M una variedad riemanniana completa, se dice que $A \subset M$ es totalmente convexo, si para cualquier par de puntos p y q de A (no necesariamente distintos), las geodésicas que unen dichos puntos, están íntegramente contenidos en A .*

Ejemplo 3.3.1 *Si $M = \mathbb{R}^n$ con la métrica identidad $G(x) = I$, cualquier conjunto convexo en el sentido clásico es totalmente convexo.*

Ejemplo 3.3.2 *Si $p \in M$ y existe una relación geodésica no trivial en p , es decir una geodésica: $\gamma : [a, b] \rightarrow M$ tal que $\gamma(a) = p = \gamma(b)$ con $\gamma(t) \neq p$ para algún $t \in [0, 1]$, entonces el conjunto $A = \{p\}$ no es totalmente convexo. Se deduce de esto que en general conjuntos unitarios no son totalmente convexos.*

Definición 3.3.2 *Decimos que $A \subset M$ 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 .*

Ejemplo 3.3.3 *El propio M y los conjuntos unitarios son conjuntos convexos.*

Definición 3.3.3 $f : M \longrightarrow \mathbb{R}$ es llamada función convexa si su restricción a cualquier geodésica de M es una función convexa en \mathbb{R} , es decir, si $\gamma : \mathbb{R} \longrightarrow M$ es una geodésica entonces:

$$f \circ \gamma : \mathbb{R} \longrightarrow \mathbb{R}$$

es convexa.

Teorema 3.3.1 $f : M \longrightarrow \mathbb{R}$ es convexa si, y solamente si, para todo segmento de geodésica $\gamma : [a, b] \longrightarrow M$ y para cualquier $\lambda \in [0, 1]$ se verifica

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

Demostración. Siendo f convexa, demostraremos que:

$$f(\gamma((1 - \lambda)a + \lambda b)) \leq (1 - \lambda)f(\gamma(a)) + \lambda f(\gamma(b)) \quad (3.6)$$

Sea $h : \mathbb{R} \longrightarrow \mathbb{R}$ tal que $h(t) = f(\gamma(t))$. Para $a, b \in [a, b]$ y $\lambda \in [0, 1]$ se tiene

$$h((1 - \lambda)a + \lambda b) \leq (1 - \lambda)h(a) + \lambda h(b).$$

De aquí se tiene (3.6).

Recíprocamente, sea $t = (1 - \lambda)a + \lambda b$ con $\lambda \in [0, 1]$ entonces:

$$\begin{aligned} f \circ \gamma(t) &= f(\gamma((1 - \lambda)a + \lambda b)) \\ &\leq f(\gamma((1 - \lambda)a)) + f(\gamma(\lambda b)) \\ &= (1 - \lambda)f(\gamma(a)) + \lambda f(\gamma(b)) \\ &\leq (1 - \lambda)f \circ \gamma(a) + \lambda f \circ \gamma(b). \quad \blacksquare \end{aligned}$$

Observación 3.3.1 La Definición 3.3.3 es la generalización natural de la definición clásica de función convexa en $M = \mathbb{R}^n$ con la métrica usual. En efecto, dados p y q la geodésica $\gamma : [0, 1] \longrightarrow M$, que los une es:

$$\gamma(\lambda) = p + \lambda(q - p) = (1 - \lambda)p + \lambda q.$$

Luego, del Teorema 3.3.1 tenemos:

$$\begin{aligned} f(\gamma(\lambda)) &= f((1 - \lambda)p + \lambda q) \leq (1 - \lambda)f(p) + \lambda f(q) \\ f((1 - \lambda)p + \lambda q) &= f(\gamma(\lambda)) = f((1 - \lambda)(0) + \lambda(1)) \leq (1 - \lambda)f(\gamma(0)) + \lambda f(\gamma(1)) \\ &= (1 - \lambda)f(p) + \lambda f(q). \end{aligned}$$

Para $\alpha \in \mathbb{R}$, definimos el conjunto de nivel $M^\alpha = \{x \in M; f(x) \leq \alpha\}$.

Teorema 3.3.2 *Si $f : M \rightarrow \mathbb{R}$ es convexa, entonces M^α es totalmente convexo.*

Demostración. Sea $p, q \in M^\alpha$ y la geodésica $\gamma : [a, b] \rightarrow M$ tal que $\gamma(a) = p$ y $\gamma(b) = q$.

Probaremos que $\gamma(t) \in M^\alpha$, para todo $t \in [a, b]$.

En efecto, sea $t = (1 - \lambda)a + \lambda b$ para algun $\lambda \in [0, 1]$, como f es convexa y por el Teorema 3.3.1 se tiene

$$\begin{aligned} f(\gamma(t)) &= f(\gamma(1 - \lambda)a + \lambda b) \leq (1 - \lambda)f(\gamma(a)) + \lambda f(\gamma(b)) \\ &= (1 - \lambda)f(p) + \lambda f(q) \\ &\leq (1 - \lambda)\alpha + \lambda\alpha = \alpha. \end{aligned}$$

Así $f(\gamma(t)) \leq \alpha$, por tanto $\gamma(t) \in M^\alpha$. ■

Teorema 3.3.3 *$f : M \rightarrow \mathbb{R}$, es convexa en p si y solo si, para cualquier geodésica $\gamma : \mathbb{R} \rightarrow M$ con $\gamma(0) = p$ vale la desigualdad*

$$f(\gamma(t)) - f(p) \geq t \langle \text{grad } f(p), \gamma'(0) \rangle. \quad (3.7)$$

Demostración. Definimos una aplicación $h : \mathbb{R} \rightarrow \mathbb{R}$ tal que $h(t) = f(\gamma(t))$, h es convexa en 0 desde que f es convexa en p y así se tiene:

$$h(t) - h(0) \geq th'(0),$$

luego,

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

Recíprocamente, si $f(\gamma(t)) - f(p) \geq \langle \text{grad } f(p), \gamma'(0) \rangle t$, esto es,

$$h(t) - h(0) \geq th'(0),$$

entonces h es convexa en 0 y por tanto f es convexa en p . ■

Teorema 3.3.4 *Si $f : M \rightarrow \mathbb{R}$ es convexa, entonces todo punto crítico de f es un punto de mínimo global de f .*

Demostración. Sea $x \in M$, debido al Teorema de Hopf-Rinow consideramos una geodésica $\gamma : \mathbb{R} \rightarrow M$ tal que $\gamma(0) = x$ y $\gamma(b) = y$, como f es convexa y del Teorema 3.3.3

$$f(\gamma(b)) - f(\gamma(0)) \geq b \langle \text{grad } f(\gamma(0)), \gamma'(0) \rangle,$$

esto es,

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

Como $\text{grad } f(x) = 0$ entonces $f(y) \geq f(x)$, para todo $y \in M$. Por tanto x es punto de mínimo global de f . ■

Teorema 3.3.5 Sea $f : M \rightarrow \mathbb{R}$ de clase C^2 , f es convexa si, y solamente si, para todo $p \in M$ la Hesiana de f en p

$$H_p^f : T_p M \rightarrow T_p M$$

es semidefnida positiva.

Demostración. Sea $v \in T_p M$ y $p \in M$ y la geodésica $\gamma : \mathbb{R} \rightarrow M$ tal que $\gamma(0) = p$ y $\gamma'(0) = v$.

Definiendo $h : \mathbb{R} \rightarrow \mathbb{R}$ tal que $h(t) = f(\gamma(t))$ sabemos que h es convexa y de clase C^2 . Del análisis convexo clásico, tenemos que esto es equivalente a $h''(t) \geq 0$, se tiene

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

y

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

Recíprocamente, si

$$\langle v, H_p^f v \rangle \geq 0,$$

entonces f es convexo. En efecto, definiendo $h : \mathbb{R} \rightarrow \mathbb{R}$ convexa, entonces se tiene $h = f \circ \gamma$ es convexa. ■

3.3.2 Funciones cuasi-convexas y pseudoconvexas

Definición 3.3.4 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 3.3.6 Sea $f : M \rightarrow \mathbb{R}$ una función diferenciable y cuasi-convexa en una variedad riemanniana completa M y sea $x, y \in M$. Si $f(x) \leq f(y)$ entonces:

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

donde $\text{grad } f$ es el gradiente de f y γ es la curva geodésica tal que $\gamma(0) = y$ y $\gamma(1) = x$.

Demostración. Sea la geodésica $\gamma : [0, 1] \rightarrow M$ tal que $\gamma(0) = y$ y $\gamma(1) = x$. Definimos $h : \mathbb{R} \rightarrow \mathbb{R}$ tal que $h(t) = f \circ \gamma(t)$, usando la aproximación de Taylor de primer orden de h en $t = 0$ tenemos:

$$h(t) = h(0) + th'(0) + \theta(t),$$

donde: $\lim_{t \rightarrow 0} \frac{\theta(t)}{t}$. Entonces tenemos:

$$f(\gamma(t)) = f(\gamma(0)) + t \langle \text{grad } f(\gamma(0)), \gamma'(0) \rangle + \theta(|t|),$$

como f es cuasi-convexa y $f(x) \leq f(y)$ tenemos:

$$t \langle \text{grad } f(y), \gamma'(0) \rangle,$$

dividiendo por t y tomando límite cuando $t \rightarrow 0$ se tiene $\langle \text{grad } f(y), \gamma'(0) \rangle \leq 0$. ■

Definición 3.3.5 Una función diferenciable $f : M \rightarrow \mathbb{R}$ es pseudoconvexa si, para todo par de puntos distintos $x, y \in M$ y toda curva geodésica que une x a y ($\gamma(0) = x$ y $\gamma(1) = y$) tenemos:

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

Teorema 3.3.7 Sea $f : M \rightarrow \mathbb{R}$ una función diferenciable y pseudoconvexa. Entonces, x^* es un mínimo global de f si, solamente si, $\text{grad } f(x^*) = 0$.

Demostración. Sea la geodésica $\gamma : \mathbb{R}_+ \rightarrow M$ tal que $\gamma(0) = x^*$ y definimos $h : \mathbb{R} \rightarrow \mathbb{R}$ con $h = f \circ \gamma$. Por el desarrollo de Taylor de primer orden de h en 0,

$$h(t) = h(0) + th'(0) + \theta(|t|),$$

donde $\lim_{t \rightarrow \infty} \frac{\theta(|t|)}{|t|} = 0$, entonces tenemos:

$$f(\gamma(t)) = f(x^*) + t \langle \text{grad } f(x^*), \gamma'(0) \rangle + \theta(|t|).$$

Como x^* es mínimo global entonces $f(x^*) \leq f(y)$, en particular para $y = \gamma(t)$ entonces,

$$f(y) - f(x^*) = t \langle \text{grad } f(x^*), \gamma'(0) \rangle + \theta(|t|),$$

luego,

$$t \langle \text{grad } f(x^*), \gamma'(0) \rangle + \theta(|t|) \geq 0,$$

que en el límite cuando $t \rightarrow 0$, $\langle \text{grad } f(x^*), \gamma'(0) \rangle \geq 0$, finalmente tomando

$$\gamma'(0) = -\text{grad } f(x^*),$$

se tiene $\text{grad } f(x^*) = 0$. El recíproco es inmediato basta usar la definición de f ser pseudoconvexa. ■

Capítulo 4

Método del Máximo Descenso

En este último capítulo estudiaremos el método de máximo descenso llamado también método del gradiente, buscaremos extender la convergencia global del método utilizando la regla de Armijo generalizado. Para este fin vamos a considerar el problema de Optimización no lineal

$$\min_{x \in M} f(x) \quad (4.1)$$

donde, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función de clase C^1 y M una variedad riemanniana completa.

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

$$x^0 \in M, \quad (4.2)$$

$$x^{k+1} = \exp_{x^k}(-t_k \text{grad} f(x^k)) \quad (4.3)$$

donde \exp_{x^k} es una aplicación exponencial en el punto x^k , t_k es un parámetro positivo, $-\text{grad} f(x)$ es el gradiente de f . En el caso de tener $M = \mathbb{R}^n$ (el espacio euclidiano) tenemos que (4.3) es equivalente a:

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

Así, el método de máximo descenso en variedades riemannianas generaliza el método clásico de máximo descenso en \mathbb{R}^n , véase idealmente sobre una superficie de \mathbb{R}^3 un esquema del funcionamiento iterativo del método del gradiente generalizado:

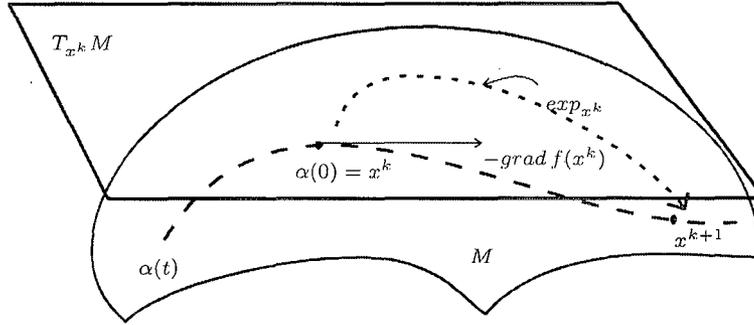


Figura 4. Esquema del proceso iterativo del método de máximo descenso sobre una variedad M .

Existen diferentes maneras de escoger el parámetro t_k generando consecuentemente diversos submétodos los cuales para su aplicación, dependerá exclusivamente de su complejidad computacional, reglas que mostramos a seguir.

Método A: Gradiente con búsqueda exacta

1. Dado x^k , calcule el $\text{grad } f(x^k)$ sobre el plano tangente $T_{x^k}M$.
2. Determine la geodésica $\gamma(t)$, $t \geq 0$, de M que verifique $\gamma(0) = x^k$ y $\gamma'(0) = -\text{grad } f(x^k)$.
3. Minimize $f(\gamma(t))$, $t \geq 0$, obteniendo t_k y defina:

$$x^{k+1} = \gamma(t_k).$$

Método B: Gradiente con Regla de Armijo

1. Dado x^k , calcule el $\text{grad } f(x^k)$ en $T_{x^k}M$.
2. Determine la geodésica $\gamma(t)$, $t \geq 0$, de M que verifique $\gamma(0) = x^k$ y $\gamma'(0) = -\text{grad } f(x^k)$.
3. Hacer:

$$t_k := 2^{-i_k},$$

donde i_k es el menor entero positivo tal que:

$$f(\gamma(t_k)) \leq f(x^k) - \alpha t_k \left\| \text{grad } f(x^k) \right\|^2$$

y $\alpha \in (0, \frac{1}{2})$.

Definición 4.0.6 Una función $f : M \rightarrow \mathbb{R}$ es llamada *gradiente Lipschitziana* con constante Γ si para todo $p, q \in M$ y $\gamma : [0, a] \rightarrow M$ la geodésica con $\gamma(0) = p$ y $\gamma(a) = q$ se verifica:

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

para todo $t \in [0, a]$, donde $P_{\gamma(t)}$ es el transporte paralelo de $\gamma(0) = p$ a $\gamma(t)$.

Método C: Gradiente con Pasos fijos

1. Dado x^k calcule el $\text{grad } f(x^k)$ en $T_{x^k}M$.
2. Determine la geodésica $\gamma(t_k)$, $t_k \geq 0$, de M que verifique $\gamma(0) = x^k$ y $\gamma'(0) = -\text{grad } f(x^k)$
3. Dados $\delta_1 > 0$ y $\delta_2 > 0$ tales que,

$$\delta_1 \Gamma + \delta_2 > 1,$$

donde Γ es la constante de Lipschitz asociada al campo gradiente de f , escoger

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

4.1 Convergencia para el caso cuasi-convexo

Estamos interesados en resolver el siguiente problema de optimización:

$$(p) \min_{x \in M} f(x)$$

donde M es una variedad riemanniana conexa, completa de dimensión finita y $f : M \rightarrow \mathbb{R}$ es una función continuamente diferenciable y cuasi-convexa.

Hipótesis A1.

El conjunto de puntos óptimos globales del problema (p), denotado por X^* , es no vacío.

Denotamos el valor óptimo de (p) por f^* . Ahora, definamos el siguiente conjunto

$$U := \{x \in M : f(x) \leq \inf_k f(x^k)\}.$$

El siguiente Lema es un resultado de gran interés el cual será usado para probar que la sucesión, generada el método de máximo descenso es cuasi-Fejér convergente a U .

Lema 4.1.1 *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. Sea $x \in U$ arbitrario. Sea también $\gamma_1 : [0, l_1] \rightarrow M$ la geodésica minimal que une x^k y x con $\gamma_1(0) = x^k$, $\|\gamma_1'(0)\| = 1$ y $\gamma_2 : [0, 1] \rightarrow M$ una geodésica que une x^k y x^{k+1} , esto es $\gamma_2(0) = x^k$, $\gamma_2(1) = x^{k+1}$ con $\gamma_2'(0) = -t_k \text{grad } f(x^k)$. Por propiedad de homogeneidad de las geodésicas, γ_2 es reparametrizada tal que

$$\gamma_2 : [0, t_k \|\text{grad } f(x^k)\|] \rightarrow M,$$

tal que $\gamma_2(t_k \|\text{grad } f(x^k)\|) = x^{k+1}$ y ahora tenemos

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

Del Teorema 2.9.2 tenemos:

$$d^2(x^{k+1}, x) \leq d^2(x^k, x) + t_k^2 \|\text{grad } f(x^k)\|^2 + 2t_k d(x^k, x) \langle \text{grad } f(x^k), \gamma_1'(0) \rangle.$$

Como f es cuasi-convexa y $f(x) \leq f(x^k)$, del Teorema 3.3.6 obtenemos que

$$\langle \text{grad } f(x^k), \gamma_1'(0) \rangle \leq 0.$$

Usando este resultado en la desigualdad anterior obtenemos el resultado deseado. ■

4.1.1 Método con búsqueda de Armijo generalizado

En 1966 LARRY ARMIJO[2], publica su trabajo: "Minimization of functions having lipschitz continuous first partial derivatives", en el que demuestra la convergencia del método del gradiente, a partir de entonces sus resultados son utilizados convenientemente en la búsqueda de mejoras o extensiones teóricas computacionales. Para

nuestro objetivo, desarrollamos el método del gradiente sobre variedades riemannianas utilizando la regla de Armijo generalizado.

En esta Subsección probamos la convergencia global de este método para el caso cuasi-convexo. Estos resultados son una generalización de KIWIEL y MURTY[12] para variedades riemannianas y extiende resultados previos de convergencia obtenidos hasta este momento, para el caso convexo por BURACHIK et al.[4] y CRUZ NETO et al.[6].

El método del máximo descenso con regla de Armijo genera una sucesión de puntos $\{x^k\}$ dados por (4.2)-(4.3) donde se satisfacen las siguientes hipótesis:

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 (4.3),

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

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

Observación 4.1.1 *Observemos que la hipótesis A2 es satisfecha por la regla de Armijo para estos valores: $\phi(t) = \alpha t$, $\beta = \alpha$, $\gamma = 2$ y $\tau_\alpha = \tau_\beta = \tau_\gamma = 1$.*

Observación 4.1.2 *La hipótesis A2 también es satisfecha por el método del gradiente con pasos fijos introducida en BURACHIK[4] y generalizada para variedades riemannianas en da CRUZ NETO[6]. En efecto, en las referencias mencionadas la regla para obtener t_k es la siguiente:*

Dados δ_1 y δ_2 tal que $\delta_1 \Gamma + \delta_2 < 1$, donde Γ es la constante de Lipschitz asociada al $\text{grad } f$, escoger

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

Definiendo $\phi(t) = \beta t^2$, con $\beta = \frac{\Gamma \delta_2}{2(1 - \delta_2)}$, $\tau_\gamma = \delta_1$, $\tau_\beta = (2/\Gamma)(1 - \delta_2)$, $\alpha \in (0, 1)$ arbitrario y $\tau_\alpha = \alpha/\beta$, garantizamos la hipótesis A2.

Proposición 4.1.1 Sea $f : M \rightarrow \mathbb{R}$ una función continuamente diferenciable y cuasi-convexa. 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. De las hipótesis A2.2 y A2.3 tenemos

$$\beta t_k^2 \|\text{grad } f(x^k)\|^2 \leq f(x^k) - f(x^{k+1}). \quad (4.4)$$

Esto implica que

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

Del Lema 4.1.1 y la Definición 1.2.8 tenemos el resultado. ■

Teorema 4.1.2 Sea $f : M \rightarrow \mathbb{R}$ una función continuamente diferenciable y cuasi-convexa. Suponga que las hipótesis A1 y A2 son satisfechas. Entonces la sucesión $\{x^k\}$ 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. De la Proposición 4.1.1, $\{x^k\}$ es cuasi-Fejér convergente en U , por tanto $\{x^k\}$ es limitado por el Teorema 1.2.1. Entonces existen \bar{x} y una subsucesión $\{x^{k_j}\}$ de $\{x^k\}$ que converge para \bar{x} . De la continuidad de f obtenemos

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

Debido a que $\{f(x^k)\}$ es una sucesión no creciente, ver (4.4), con una subsucesión que converge para $f(\bar{x})$, toda la sucesión converge para $f(\bar{x})$ y así

$$f(\bar{x}) \leq f(x^k), \text{ para todo } k \in \mathbb{N}.$$

Esto implica que $\bar{x} \in U$. Ahora, del Teorema 1.2.1, concluimos que $\{x^k\}$ converge para \bar{x} . Finalmente, probaremos que $\text{grad } f(\bar{x}) = 0$. Por contradicción, supongamos que $\text{grad } f(\bar{x}) \neq 0$.

Claramente, tenemos que $\text{grad } f(x^k) \rightarrow \text{grad } f(\bar{x}) \neq 0$ y $f(x^k) \rightarrow f(\bar{x})$. Ahora, de (4.4), se cumple que

$$\lim_{k \rightarrow +\infty} t_k = 0. \quad (4.5)$$

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

$$f(\exp_{x^k}(-\bar{t}_k \text{grad } f(x^k))) - f(x^k) \geq -\alpha \bar{t}_k \|\text{grad } f(x^k)\|^2. \quad (4.6)$$

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

$$-\langle \text{grad } f(\exp_{x^k}(-t_k^* \text{grad } f(x^k))), P_{\gamma_k, 0, t_k^*} \text{grad } f(x^k) \rangle \geq -\alpha \|\text{grad } f(x^k)\|^2,$$

donde $P_{\gamma_k, 0, t_k^*}$ es el transporte paralelo a lo largo de la geodésica γ_k tal que $\gamma_k(0) = x^k$ y $\gamma_k'(0) = -\text{grad } f(x^k)$. Ahora, (4.5) y A2.4 implican que $\lim_{k \rightarrow +\infty} t_k^* = 0$. Haciendo $k \rightarrow +\infty$ en la desigualdad anterior y tomando en cuenta la continuidad de $\text{grad } f$, \exp y el transporte paralelo, tenemos que $1 \leq \alpha$, lo que contradice A2.1. Por tanto, $\text{grad } f(\bar{x}) = 0$. ■

Como consecuencia inmediata del teorema anterior y del Teorema 3.3.7 tenemos el siguiente resultado.

Corolario 4.1.1 *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 (p).*

Ejemplo 4.1.1 *Sea la función $f(x_1, x_2) = (\ln x_1)^2 + (\ln x_2)^2$. cuya simulación se ve en las figuras 5, 6 y 7 a escala multiplicada por 10.*

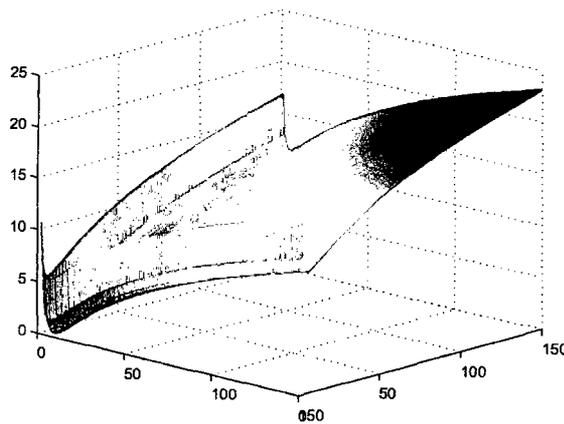


Figura 5. Gráfico de la función f .

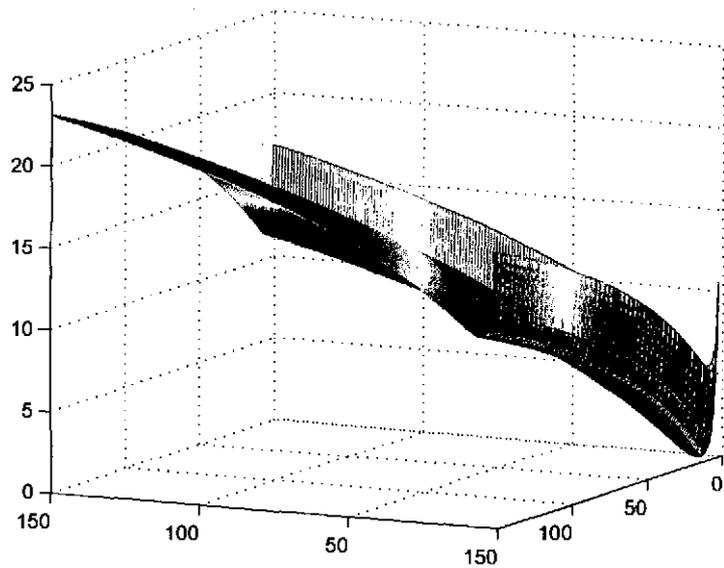


Figura 6. Gráfico de la función f después de una rotación.

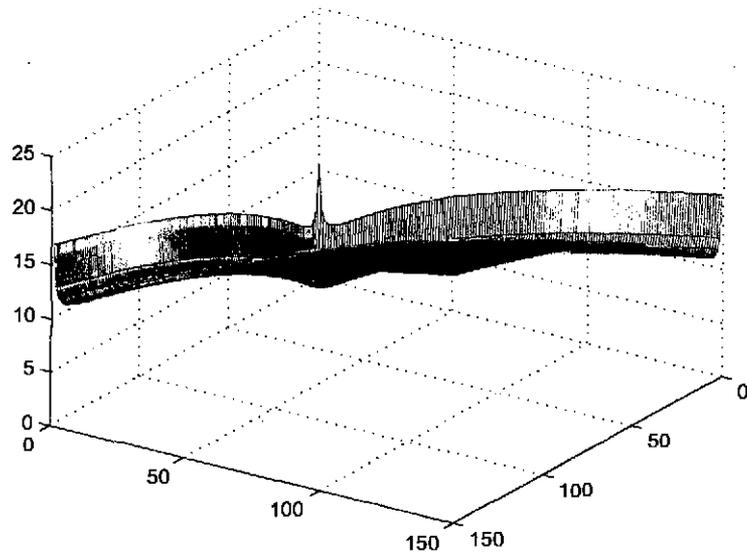


Figura 7. Vista frontal del gráfico de f

Esta función f es claramente no convexa en \mathbb{R}^2 con la métrica usual, sin embargo al hacer uso de la métrica $G(p) = \text{diag}(1/(p_i)^2)$ con $i = 1, 2.$, la función en cuestión se

Esta función f es claramente no convexa en \mathbb{R}^2 con la métrica usual, sin embargo al hacer uso de la métrica $G(p) = \text{diag}(1/(p_i)^2)$ con $i = 1, 2$, la función en cuestión se transforma en convexa sobre la variedad \mathbb{R}_{++}^2 , dado que su Hessiano es semidefinida positiva, a saber,

$$H_x^f = \begin{bmatrix} \frac{2}{x_1} & 0 \\ 0 & \frac{2}{x_2} \end{bmatrix}.$$

Su ecuación geodésica es

$$\gamma(t) = p_i \exp^{\frac{v_i}{p_i} t}, \quad i = 1, 2$$

y las iteraciones del método de máximo descenso son:

$$x_i^{k+1} = x_i^k \exp\left(t_k x_i^k \frac{\partial f(x^k)}{\partial x_i}\right), \quad i = 1, 2.$$

Discusiones

En este trabajo la hipótesis de curvatura no negativa es esencial para obtener la convergencia global del método a un punto crítico para funciones cuasi-convexas continuamente diferenciables. Así, para espacios donde la curvatura es negativa, como por ejemplo los espacios de Lobachevsky, no sabemos si nuestro algoritmo converge o no (globalmente). Esto nos lleva a formular la siguiente interrogante: ¿Podrá obtenerse la convergencia global del método sin usar hipótesis sobre la curvatura de la variedad riemanniana?

Conclusiones

1. El resultado más importante de esta Tesis, es que extendemos la teoría de convergencia global del método de máximo descenso para la clase de funciones cuasi-convexas definidas sobre variedades riemannianas de curvatura seccional no negativa. Así el campo de aplicaciones del método es ampliado para diversos problemas de optimización no convexa donde sus restricciones pueden ser vistos como una variedad riemanniana completa de curvatura seccional no negativa y la función objetivo sea cuasi-convexa en esta variedad.
2. Mediante la introducción de una métrica diagonal riemanniana sobre el ortante positivo \mathbb{R}_{++}^n damos caracterizaciones para el gradiente y Hessiana de una función en esta variedad. En particular, presentamos condiciones para que funciones no convexas en el sentido usual se transformen en convexas en la variedad riemanniana \mathbb{R}_{++}^n con dicha métrica, mostramos esta afirmación con un ejemplo novedoso.
3. La interrogante surgida en las Discusiones, motiva a un estudio más profundo del método sin usar hipótesis sobre la curvatura de la variedad riemanniana.

Bibliografía

- [1] ALEKSANDROV A. D., KOLMOGOROV A. N. and LAURENTIEV M. A., *La Matemática: su contenido, métodos y significado*. Edit. Alianza Universidad, Madrid-España, 1981.
- [2] ARMIJO L., “*Minimization of functions having lipschitz continuous first partial derivatives*”. Pacific Journal of Mathematics, Vol. 16, No. 1, 1966.
- [3] BOOTHBY W. M. *An Introduction to differentiable manifold and riemannian geometry*. Edit. Academic Press. Orlando, Florida USA., 1986.
- [4] BURACHIK R. S., GRANA DRUMOND L. M., IUSEM A. N and SVAITER B. F., “*Full convergence of the steepest descent method with inexact line searches*”. Optimization, 32 (1995), 137-145.
- [5] da CRUZ NETO, J. X, OLIVEIRA, P. R., “*Geodesic methods in riemannian manifolds*”, Technical Report ES-352/95, Systems Engineering and Computer Sciences, PESC/COPPE, Federal University of Rio de Janeiro, 1995.
- [6] da CRUZ NETO, J. X., de LIMA, L. L., OLIVEIRA, P. R., “*Geodesic algorithms in riemannian geometry*”. Balkan Journal of Geometry and its Applications BJGA. 1998, v. 3, n.2, pp. 89-100.
- [7] do CARMO, M. P., *Geometria Diferencial de Curvas e Superficies*, Sociedade Brasileira de Matemática, Rio de Janeiro, Brasil, 2005.
- [8] do CARMO, M. P., *Geometria Riemanniana*, 2 ed. IMPA-Projeto Euclides, Rio de Janeiro, Brasil, 1988.

- [9] FERREIRA O. P. and OLIVEIRA P. R., “Subgradient algorithm on riemannian manifold”, Journal of Optimization. Theory and Applications, 97(1) (1998), 93-104.
- [10] GABAY, D., “Minimizing a differentiable function over a differentiable manifold”. Journal of Optimization Theory and Application, 1982, v. 37, pp., 177-219.
- [11] HICKS N. J. *Notes on differential geometry*. Van Nostrand Reinhold Company. The University of Michigan-USA, 1966.
- [12] KIWIEL K. C., MURTY K., “Convergence of the steepest descent method for minimization quasiconvex functions”. JOTA, Journal of Optimization. Theory and Applications, 89 (1), 1996, 221-223.
- [13] LAGES L. E. *Introducao às variedades diferenciáveis*. EMMA. Porto Alegre - Rio de Janeiro, Brasil, 1960.
- [14] LAGES L. E. *Variedades diferenciáveis*. IMPA, Rio de Janeiro-Brasil, 1973.
- [15] LUENBERGER D. G., “The gradient projection method along geodesics”, Management Science, 1972(18), n. 11, pp. 620-631.
- [16] LUENBERGER D. G., *Introduction to linear and nonlinear programming*, 2 ed. Addison-Wesley, Massachusetts, USA., 1973.
- [17] OLIVEIRA P. R., CRUZ NETO, J. X., “Elementos de geometria riemanniana”, Technical Report, Systems Engineering and Computer Science, PESC/COPPE, Federal University of Rio de Janeiro, 1995.
- [18] PAPA QUIROZ E. A. and OLIVEIRA P. R., “A new self-concordant barrier for the hypercube”. JOTA, Journal of Optimization Theory and Applications, 2007, 135(3), 1-10.
- [19] PAPA QUIROZ E. A., QUISPE E. M. and OLIVEIRA P. R. “Steepest descent method with a generalized Armijo search for quasiconvex functions on riemannian manifolds”. J. Math. Anal. Appl. 341(2008), 467-477.

- [20] RAPCSÁK T., *Smooth nonlinear optimization in \mathbb{R}^n* , Academic Publishers, Kluwer USA., 1997.
- [21] SMITH S.T., "*Optimization techniques on riemannian manifolds*", Fields Institute Communications, AMS, Providence, RI, 1994(3), 113-146.
- [22] UDRISTE C. *Convex functions and optimization methods on riemannian manifolds*, Academic Publishers, Kluwer USA., 1997.
- [23] WRIGHT S. J., *Primal-dual interior point methods*. 1 ed. Philadelphia, SIAM, USA., 1997.