

UNIVERSIDAD CENTROCCIDENTAL “LISANDRO ALVARADO”
DECANATO DE CIENCIAS Y TECNOLOGÍA
LICENCIATURA EN CIENCIAS MATEMÁTICAS



Método de Cauchy con Búsqueda generalizada de Armijo para funciones casi-convexas en Variedades Riemannianas

AUTOR: BR. LAURA JOSEFINA VALLADARES FRASER
TUTOR: DR. EIBAR HERNÁNDEZ (UCLA)

TRABAJO ESPECIAL DE GRADO

Presentado ante la Ilustre
Universidad Centroccidental “Lisandro Alvarado”
como requisito final para optar al grado de
Licenciado en Ciencias Matemáticas.

BARQUISIMETO, VENEZUELA

Febrero, 2012.

A *Dios* todopoderoso
Mi Madre, Mi Tía,
Mis abuelas...

Agradecimientos

Primeramente a *Dios*, por guiarme en el camino correcto para lograr cada meta que me he propuesto, esta en especial, por hacer de mi una persona fuerte, madura, iluminar mi mente, y por colocar en mi camino a todas aquellas personas que han sido mi soporte y compañía durante todo el periodo de estudio.

A *Mi Madre y mi tia*, por ser madres ejemplares, por su apoyo incondicional en cada momento de mi vida, por darme la estabilidad emocional y económica para llegar a este logro, sin su apoyo nada de esto hubiese sido posible, las amo.

A *mis abuelas*, por ser pilar fundamental en mi vida, las adoro.

A *mi tío Rafael*, por toda su ayuda, apoyo y consejos de padre durante toda mi vida, lo adoro.

A *Edward Santeliz*, por ese *Amor* que me da día a día, por todo su apoyo, confianza, consejos y regaños en los momentos que mas lo necesite, gracias gordito.

A mi *tutor* Eibar Hernández, por su dedicación, por ayudarme en la realización de éste proyecto, por explicarme y transmitirme de manera paciente sus conocimientos y por ser más que mi tutor un buen amigo y consejero.

A mis *Amigos* Maria José, Piero y Freddy por ser mi factor distractor durante toda mi carrera, gracias por su apoyo, los quierooooo.

A mis *Amigos* Ismerai, Yorgeth, Carlitos, Rafael, Yori, Naty, Yese, Ifi, Kimberly por todo su apoyo a lo largo de toda la carrera, sin su amistad no hubiese logrado adaptarme a barquisimeto, los quierooooo, muchisimas gracias.

A *mis tias, tios, primas, primos* y todas aquellas personas que de una u otra forma colocaron un granito de arena para contribuir a que se culminara esta meta.

Introducción

El Método de Cauchy , es uno de los métodos más antiguo y reconocidos en la literatura para resolver problemas de optimización. Sin embargo, los resultados clásicos de una función objetivo arbitraria, no son muy fuertes debido a que la convergencia plena y la existencia de puntos clausura no están asegurados. Solo se puede asegurar que cualquier punto clausura, si existe, es un punto crítico del problema.

Por otro lado, la generalización de los espacios euclidianos a variedades Riemanniana tienen algunas ventajas importantes. Por ejemplo, problemas de optimización con restricciones se pueden ver sin restricciones desde el punto de vista de la geometría Riemanniana, otra ventaja es que los problemas de optimización con funciones objetivos no-convexas, se pueden convertir en convexas a través de la introducción de una adecuada métrica Riemanniana.

La motivación para estudiar este tema proviene de la amplia gama de aplicaciones de optimización casi-convexa en diversas áreas de las ciencias y la ingeniería, en este contexto, si las restricciones de un problema de minimización constituyen una variedad Riemanniana, y la función objetivo es casi-convexo, el problema se convierte en un problema sin restricciones, y por lo tanto no es necesario hacer proyecciones en cada iteración del Método de Cauchy.

El propósito de este trabajo es estudiar y desarrollar el artículo [1], donde se analiza y generaliza el Método de Cauchy para el caso casi-convexo en

variedades Riemannianas con curvatura seccional no-negativa. Se prueba la convergencia plena de este método a un punto crítico del problema con una búsqueda generalizada De Armijo y una regularización proximal, utilizando la teoría de convergencia casi-Fejér.

Nuestro interés es resolver problemas de optimización,

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

donde M es una variedad Riemanniana completa finito dimensional y $f : M \rightarrow \mathbb{R}$ una función casi-convexa continuamente diferenciable.

El método de Cauchy para Variedades Riemannianas para el caso casi-convexo es el siguiente:

El método de Cauchy genera una sucesión $\{x^k\}$,

$$x^0 \in M \tag{1}$$

$$x^{k+1} = \exp_{x^k}(-t_k \text{grad}f(x^k)) \tag{2}$$

donde \exp es la exponencial y t_k es algún tamaño de paso positivo.

El método de Cauchy con búsqueda generalizada de Armijo para Variedades Riemannianas para el caso casi-convexo es el siguiente:

El Método de Cauchy con tamaño de paso de Armijo genera una sucesión $\{x^k\}$ dada por (1) y (2) donde:

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

El método de Cauchy con una Regularización Proximal en Variedades Riemannianas para el caso casi-convexo es el siguiente:

Sea $\{\lambda_k\}$ una sucesión tal que $\lambda' \leq \lambda_k \leq \lambda''$, donde $0 \leq \lambda' \leq \lambda''$. La regularización del método de Cauchy genera una sucesión $\{x^k\}$ definida por (2.2) y (2.3), tal que:

$$t_k = \operatorname{argmin}_{t>0} \varphi_k(t), \tag{3}$$

donde

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

y sea;

$$x^{k+1} = \exp_{x^k}(-t_k \operatorname{grad}f(x^k)) \tag{4}$$

En el capítulo uno, se mencionan algunos preliminares necesarios para entender el trabajo, conceptos y teoremas básicos tanto para el espacio Euclídeo como en Variedades Riemannianas, el Método de Cauchy en \mathbb{R}^n , búsqueda de Armijo y la Regularización proximal. En el capítulo dos, capítulo de desarrollo, se analiza el Método de Cauchy para el caso casi-convexo en Variedades Riemanniana con curvatura seccional no-negativa, se prueba la convergencia plena de este método a un punto crítico del problema utilizando una búsqueda generalizada de Armijo y una regularización proximal.

Índice general

Introducción	III
1. Preliminares	1
1.1. Conceptos Básicos	1
1.2. Método de Cauchy en \mathbb{R}^n y Regularización Proximal	3
1.2.1. Método de Cauchy (Direcciones de descenso, reglas de medidas de paso)	3
1.2.2. Regularización Proximal	7
1.3. Algunos resultados Preliminares	9
1.4. Superficies Diferenciables	10
1.5. La primera forma fundamental	12
1.6. Geometría de la Aplicación de Gauss	17

1.7. Superficies Abstractas y Variedades Riemannianas.	20
1.8. Transporte Paralelo y Geodésica	25
1.9. La Función Exponencial.	31
2. Desarrollo	38
2.1. Método de Cauchy	38
2.2. Método de Cauchy con búsqueda generalizada de Armijo	42
2.3. El Método de Cauchy con una regularización Proximal, para el caso casi-convexo.	46
Conclusión	I

Índice de figuras

1.1. Métrica Riemanniana	13
1.2. Variedad Diferencial.	21
1.3. Plano Tangente.	22
1.4. Derivada Covariante	26
1.5. Campo Vectorial Paralelo	27
1.6. Transporte Paralelo	27
1.7. Geodésica	29
1.8. Función Exponencial.	32
1.9. Teorema de Hopf-Rinow.	32
1.10. Bisagra Geodesica.	34
1.11. Teorema de Toponogov.	35

2.1. Esquema del proceso iterativo del método de Cauchy sobre una variedad M	40
---------------------------------------------------------------------------------------------	----

Capítulo 1

Preliminares

1.1. Conceptos Básicos

Se presentarán algunas definiciones del análisis convexo que se necesitaran para una mejor comprensión de este trabajo. A lo largo de éste, \mathbb{R} denotará el sistema de números reales y \mathbb{R}^n el espacio vectorial usual de n-uplas de la forma $x = (x_1, \dots, x_n)^t$, donde cada $x_i \in \mathbb{R}$, para $i = 1, \dots, n$ y donde cada x^t denotará el vector traspuesto de x .

El producto interno de dos vectores $x, y \in \mathbb{R}^n$ será representado por:

$$\langle x, y \rangle = x_1y_1 + \dots + x_ny_n = x^ty.$$

Definición 1.1. *Un subconjunto C de \mathbb{R}^n es llamado convexo si y sólo si*

$$(1 - \lambda)x + \lambda y \in C,$$

para cada $x, y \in C, \lambda \in (0, 1)$.

Es decir, un conjunto es convexo, cuando el segmento de recta que une a cualquier par de puntos del conjunto está totalmente contenido en éste.

Definición 1.2. Sea C un conjunto convexo y no-vacío en \mathbb{R}^n . Una función $f : C \rightarrow \mathbb{R}$ se dice convexa sobre C si y sólo si para cada par $(x, y) \in C \times C$ y para cada $\lambda \in (0, 1)$ se tiene que:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Se dice que f es estrictamente convexa sobre C , cuando la desigualdad anterior es estricta para $x \neq y$.

Definición 1.3. Una función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ no idénticamente igual a $+\infty$ se dice que es convexa, cuando para todo par $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ y para todo $\lambda \in (0, 1)$ se cumple que:

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Obsérvese que la desigualdad anterior es en $\mathbb{R} \cup \{+\infty\}$ y denotaremos a la clase de tales funciones por $\text{Conv}\mathbb{R}^n$.

Definición 1.4. Dada una función $f : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ no idénticamente igual a $+\infty$, el epígrafo de f es el conjunto no-vacío dado por:

$$\text{epi}(f) := \{(x, r) \in \mathbb{R}^n \times \mathbb{R} : r \geq f(x)\}.$$

El epígrafo estricto es definido de manera similar, reemplazando " \geq " por " $>$ ".

Definición 1.5. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$ una función diferenciable, diremos que $x \in \mathbb{R}^n$ es punto crítico si se verifica que:

$$\nabla f(x) = 0.$$

Definición 1.6. Sea $f : C \rightarrow \mathbb{R}$ siendo C un conjunto convexo. Diremos que f es *casiconvexa*, si para todo $x, y \in C$ se cumple que:

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}, \text{ con } \lambda \in (0, 1).$$

1.2. Método de Cauchy en \mathbb{R}^n y Regularización Proximal

1.2.1. Método de Cauchy (Direcciones de descenso, reglas de medidas de paso)

Cuando del área de optimalidad se trata, la principal idea de métodos de optimización irrestricto tiene una fácil explicación geométrica, pero el análisis de convergencia que le corresponde es con frecuencia complejo.

Se considera el problema de minimización irrestricta de una función diferenciable continua $f : \mathbb{R}^n \rightarrow \mathbb{R}$. El algoritmo mas interesante de este problema se desarrollará en una idea confiable, llamada descenso iterativo, este procedimiento consiste en lo siguiente:

Se comenzará con un punto x_0 y una sucesión de vectores generada $\{x^k\}$ tal que f es decreciente en cada iteración, esto es:

$$f(x_{k+1}) < f(x_k); k = 0, 1, \dots$$

y haciéndolo de esta forma, mejoramos sucesivamente nuestra solución estimada, esperando encontrar el mínimo por esta via. En el descenso iterativo para minimización de una función f , cada vector generado por la sucesión tiene costo menor que el de su predecesor.

Dado un vector $x_0 \in \mathbb{R}^n$ con $\nabla f(x_0) \neq 0$ considere la semi recta de vectores,

$$x_{k+1} = x_k + \alpha_k \nabla f(x_k), \forall \alpha_k \geq 0.$$

De la expansión de la serie de Taylor de primer orden alrededor de x_{k+1} y usando la ecuación anterior tenemos:

$$f(x_{k+1}) = f(x_k) + (\nabla f(x_k))^t (x_{k+1} - x_k) + o(\|x_{k+1} - x_k\|) = f(x_k) - \alpha_k \|\nabla f(x_k)\|^2 + o(\alpha_k \|\nabla f(x_k)\|).$$

donde $o(\alpha_k \|\nabla f(x_k)\|)$ es una función $\rho(\alpha_k)$ tal que $\lim_{\alpha_k \rightarrow 0} \frac{\rho(\alpha_k)}{\alpha_k \|\nabla f(x_k)\|} = 0$.

Esto es, cuando el termino α_k es suficientemente pequeño, se obtendrá lo siguiente:

$$\begin{aligned} f(x_{k+1}) &\simeq |f(x_k) - \alpha_k \|\nabla f(x_k)\|^2| \\ &\leq f(x_k). \end{aligned}$$

Si $\nabla f(x_0) \neq 0$ se repite el procedimiento, pero si $\nabla f(x_0) = 0$ entonces x_{k+1} es el minimizador de la función f .

Llevando esta idea un paso mas lejos, consideremos:

$$x_{k+1} = x_k - \alpha_k d_k, \forall \alpha_k \geq 0$$

donde la iteración del vector $d_k \in \mathbb{R}^n$ le hacemos un ángulo mayor de noventa grados con el gradiente, esto es:

$$(\nabla f(x_k))^t d_k < 0.$$

Luego por el teorema de Taylor tenemos:

$$f(x_{k+1}) = f(x_k) - \alpha_k \nabla f(x_k)^t d_k + o(\alpha_k).$$

ocurriendo lo mismo que en el caso anterior para α_k cercano a cero. Las observaciones anteriores forman la base del algoritmo amplio e importante,

$$x_{k+1} = x_k - \alpha_k d_k, k = 0, 1, 2, \dots$$

donde $\nabla f(x_0) \neq 0$ y la dirección d_k es tomada tal que:

$$(\nabla f(x^k))^t d^k < 0.$$

y el tamaño de paso α_k es tomado positivo. Si $\nabla f(x_0) = 0$ el método se detiene, esto es, $x_{k+1} = x_k$, lo cual es equivalente a escoger $d_k = 0$.

En vista de la relación $(\nabla f(x^k))^t d^k < 0$ para la dirección d^k y el gradiente $\nabla f(x^k)$ los algoritmos de este tipo son llamados métodos gradientes. La gran mayoría de los métodos gradientes que consideramos son también algoritmos de descenso, esto es, el tamaño de paso α^k es seleccionado tal que:

$$f(x^k + \alpha^k d^k) < f(x^k), k = 0, 1, 2, \dots$$

Sin embargo, estos son algunas excepciones. Podemos hablar de muchas posibilidades a la hora de escoger la dirección d^k y el tamaño de paso α^k en un método de descenso, efectivamente el método gradiente, no es el único que cumple con esta función.

En este trabajo se estudiará y desarrollará un análisis de convergencia para una situación particular en el método de Cauchy.

La forma general del algoritmo bajo consideraciones es:

$$x^0 \in \mathbb{R}^n \tag{1.1}$$

$$x^{k+1} = x^k - \lambda_k \nabla f(x^k) \tag{1.2}$$

donde el valor de $\lambda_k > 0$ varía según el criterio.

Definición 1.7. Una sucesión $\{y^k\}$ es casi-Fejér convergente a $U \subset \mathbb{R}^n$ si para todo $u \in U$, existe ϵ_k sucesión de \mathbb{R} tal que $\epsilon_k \geq 0, \forall k, \sum_{k=0}^{\infty} \epsilon_k < \infty$ y

$$\|y^{k+1} - u\|^2 \leq \|y^k - u\|^2 + \epsilon_k, \forall k.$$

Proposición 1.1. Si $\{y^k\}$ es casi-Fejér convergente a un conjunto no vacío $U \subset \mathbb{R}^n$, $\{y^k\}$ es acotada. Si además y es un punto clausura de $\{y^k\}$ perteneciente a U , entonces $\lim_{k \rightarrow \infty} y^k = y$.

Demostración:

Probemos primero que $\{y^k\}$ es acotada. Sea $u \in U$, por la definición anterior tenemos que:

$$\begin{aligned} \|y^k - u\|^2 &= \|y^{k-1} - u\|^2 + \epsilon_{k-1} \\ &\leq \|y^{k-2} - u\|^2 + \epsilon_{k-2} + \epsilon_{k-1} \\ &\cdot \\ &\cdot \\ &\cdot \\ &\leq \|y^{k-k} - u\|^2 + \epsilon_0 + \epsilon_1 + \dots + \epsilon_{k-2} + \epsilon_{k-1} \\ &= \|y^0 - u\|^2 + \sum_{j=0}^{k-1} \epsilon_j \\ &= \|y^0 - u\|^2 + \sum_{j=0}^{\infty} \epsilon_j. \end{aligned}$$

Así, fijando u , tenemos que $\{y^k\}$ es acotada ya que esta contenida dentro de una bola con centro en u y radio $\|y^0 - u\|^2 + \sum_{j=0}^{\infty} \epsilon_j$.

Sea $y \in U$ un punto clausura de $\{y^k\}$ y $\delta > 0$ entonces existe $\{y^{l_k}\}$ sub-sucesión de $\{y^k\}$ tal que $y^{l_k} \rightarrow y$, como $\{y^k\}$ es casi-Fejér, existe $\{\epsilon_k\} \subset \mathbb{R}$ tal

que $\epsilon_k \geq 0$ y $\sum_{j=0}^{\infty} \epsilon_j < \infty$. Para $\delta > 0$ existe k_0 tal que $\sum_{j=0}^{\infty} \epsilon_j < \frac{\delta}{2}$ y existe $k_1 \geq k_0$ tal que: $k \geq k_1 \Rightarrow \|y^{l_k} - y\|^2 < \frac{\delta}{2}$, así

$$k \geq k_1 \Rightarrow \|y^k - y\|^2 \leq \|y^{l_k} - y\|^2 + \sum_{j=0}^{\infty} \epsilon_j < \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

Por tanto, $\lim_{k \rightarrow \infty} y^k = y$. □

1.2.2. Regularización Proximal

La idea de regularización surge en conexión con un problema que no se comporta bien en el siguiente sentido. Dado un problema de forma:

$$L(f) = 0. \tag{1.3}$$

donde f es un elemento de un conjunto X , por ejemplo un espacio de funciones y $L : X \rightarrow X$ es un operador (usualmente diferenciable). Se dice que "mal puesto" cuando no tiene solución o tiene más de una solución o tiene solución pero esta no depende de manera continua respecto a los parámetros de L .

El trabajo consiste en reemplazar L por un operador regularizable $L + \lambda M$ ($\lambda \in \mathbb{R}$, $M : X \rightarrow X$), donde M es tal que el problema:

$$L(f) + \lambda M(f) = (L + \lambda M)(f) = 0, \tag{1.4}$$

sea bien comportada para $\lambda > 0$, en tal caso (1.4) tiene solución única f_λ , y se espera que para λ cercano a 0, f_λ , nos de una solución aproximada de (1.3).

Veamos este concepto aplicado a un problema de optimización tomando $X = \mathbb{R}^n$ y $L = \nabla f$, donde $f : \mathbb{R}^n \rightarrow \mathbb{R}$ es una función convexa. En este caso nos queda:

$$\nabla f(x) = 0. \tag{1.5}$$

Condición utilizada para resolver:

$$\min_{x \in \mathbb{R}^n} f(x). \quad (1.6)$$

Se supondrá que f es acotada inferiormente y se tomara $g : \mathbb{R} \rightarrow \mathbb{R}$ estrictamente convexa y coerciva (esto es, $g(x) \rightarrow +\infty$ cuando $\|x\| \rightarrow +\infty$) el problema (1.5) puede no tener solución, pero el problema regularizado,

$$\min_{x \in \mathbb{R}^n} \{f(x) + \lambda g(x)\}. \quad (1.7)$$

tiene una única solución para $\lambda > 0$ ya que $f + \lambda g$ es coerciva (usando el hecho de que f es acotada inferiormente), el cual reduce el problema a conjuntos compactos, garantizando la existencia de soluciones y por ser estrictamente convexa, se obtendrá la unicidad de la solución.

El problema regularizado (1.7) tiene única solución $x(\lambda)$ y bajo algunas hipótesis razonables incluyendo la existencia de la solución de (1.6), puede ser probado que el límite cuando λ tiende a cero de $x(\lambda)$ existe y resuelve (1.6).

El detalle acerca de esta regularización es que aunque $f + \lambda g$ es estrictamente convexa y coerciva para algún $\lambda > 0$ pequeño, esta función puede comportarse casi como f o en otras palabras si el sistema $\nabla f(x) = 0$ no se comporta bien, entonces el sistema $(\nabla f + \lambda \nabla g)(x) = 0$ tampoco se comportará bien cuando λ se aproxime a cero, a pesar del hecho que tiene una única solución para todo $\lambda > 0$.

Con la intención de sobrellevar la dificultad antes mencionada presentamos el algoritmo de punto proximal para optimización en \mathbb{R}^n de la siguiente forma:

$$x^0 \in \mathbb{R}^n. \quad (1.8)$$

$$x^{k+1} = \operatorname{argmin}_{x \in \mathbb{R}^n} \{f(x) + \lambda_k \|x - x^k\|^2\}. \quad (1.9)$$

donde λ_k es un numero real satisfaciendo:

$$0 < \lambda_k \leq \tilde{\lambda}$$

para $\tilde{\lambda} > 0$ (incluyendo el caso para λ_k constante) y $\|\cdot\|$ es la norma Euclidiana.

1.3. Algunos resultados Preliminares

Definición 1.8. Sea (\mathbb{X}, d) un espacio Métrico completo. La sucesión $\{y^k\}$, $k \geq 0$, \mathbb{X} es casi-Féjer convergente en un conjunto $U \subset \mathbb{X}$, si para cualquier $u \in U$, existe una sucesión $\{\epsilon_k\} \subset \mathbb{R}$ 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 1.1. En un espacio Métrico completo (\mathbb{X}, d) , si $\{y^k\}$ es casi-Fejér convergente a un conjunto no vacío $U \subset \mathbb{X}$, entonces $\{y^k\}$ es acotada. Si además \bar{y} es un punto de acumulación de $\{y^k\}$ perteneciente a U , entonces $\{y^k\}$ converge y $\lim_{k \rightarrow \infty} y^k = \bar{y}$.

Demostración:

De la definición de casi-Féjer convergente, probemos que $\{y^k\}$ es acotada. Sea $u \in U$,

$$\begin{aligned}
 d^2(y^k - u) &= d^2(y^{k-1} - u) + \epsilon_{k-1} \\
 &\leq d^2(y^{k-2} - u) + \epsilon_{k-2} + \epsilon_{k-1} \\
 &\cdot \\
 &\cdot \\
 &\cdot \\
 &\leq d^2(y^{k-k} - u) + \epsilon_0 + \epsilon_1 + \dots + \epsilon_{k-2} + \epsilon_{k-1} \\
 &= d^2(y^0 - u) + \sum_{j=0}^{k-1} \epsilon_j \\
 &= d^2(y^0 - u) + \sum_{j=0}^{\infty} \epsilon_j.
 \end{aligned}$$

Así, fijando u , tenemos que $\{y^k\}$ es acotada ya que esta contenida dentro de una bola con centro en u y radio $d^2(y^0 - u) + \sum_{j=0}^{\infty} \epsilon_j$.

Sea $y \in U$ un punto de acumulación de $\{y^k\}$ y $\delta > 0$ entonces existe $\{y^{l_k}\}$ subsucesión de $\{y^k\}$ tal que $y^{l_k} \rightarrow y$, como $\{y^k\}$ es casi-Fejér, existe $\{\epsilon_k\} \subset \mathbb{R}$ tal que $\epsilon_k \geq 0$ y $\sum_{j=0}^{\infty} \epsilon_j < \infty$. Para $\delta > 0$ existe k_0 tal que $\sum_{j=0}^{\infty} \epsilon_j < \frac{\delta}{2}$ y existe $k_1 \geq k_0$ tal que: $k \geq k_1 \Rightarrow d^2(y^{l_k} - y) < \frac{\delta}{2}$.

Así,

$$k \geq k_1 \Rightarrow d^2(y^k - y) \leq d^2(y^{l_k} - y) + \sum_{j=0}^{\infty} \epsilon_j < \frac{\delta}{2} + \frac{\delta}{2} = \delta.$$

Por tanto, $\lim_{k \rightarrow \infty} y^k = y$. □

1.4. Superficies Diferenciables

Definición 1.9. Sea $f : U \subset \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$ diferenciable en U , diremos que c es un valor regular de f cuando $\forall x \in U : f(x) = c$ (es decir, $\forall x \in f^{-1}(c)$, esto

es, $f'(x) : \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$ es sobreyectiva, $[f']$ tiene rango n).

Teorema 1.2. Sea $c \in \mathbb{R}^n$ un valor regular de $f : U \subset \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$ de clase C^k entonces $M = f^{-1}(c) = \{x \in U / f(x) = c\}$ es una superficie C^k con $\dim M = m$; $M \subset \mathbb{R}^{m+n}$ y $T_p M = \text{Ker } f'(p), \forall p \in M$.

Demostración:

Por teorema de la función inversa se tiene que $M = f^{-1}(c)$ es localmente el gráfico de una función g de clase C^k , entonces M es superficie parametrizada por $\varphi(x) = (x, g(x)), x \in \mathbb{R}^m$ así $\dim M = m$.

Ahora probemos que $\forall p \in M, T_p M = \text{Ker } f'(p)$.

Sea $p \in M$ y $v \in T_p M$, así $v = \lambda'(0)$ para alguna $\lambda : (-\varepsilon, \varepsilon) \rightarrow M$ con $\lambda(0) = p, \lambda$ diferenciable. Así,

$$f'(p)(v) = df_p(v) = (f \circ \lambda)'(0) = 0.$$

$$\begin{aligned} \lambda(t) \in M &\Rightarrow f(\lambda(t)) = c \Rightarrow (f \circ \lambda)(t) = c \\ &\Rightarrow (f \circ \lambda)'(t) = 0, \forall t \in (-\varepsilon, \varepsilon). \end{aligned}$$

Luego,

$$f'(p)(v) = 0 \Rightarrow v \in \text{Ker } f'(p).$$

Por tanto $T_p M \subset \text{Ker } f'(p)$.

Por otro lado, tenemos que:

$$\dim \text{Ker } f'(p) + \dim \text{Im } f'(p) = m + n \Rightarrow \dim \text{Ker } f'(p) = m.$$

Pero $\dim T_p M = m$ y como $T_p M \subset \text{Ker } f'(p)$ entonces podemos concluir que:

$$T_p M = \text{Ker } f'(p).$$

□

Observación 1.1.

- Sea $M \subset \mathbb{R}^3$ una superficie en \mathbb{R}^3 , donde tenemos un producto interno $\langle \cdot, \cdot \rangle$ definido como:

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + x_3 y_3.$$

- Consideremos $T_p M$, si tenemos $w_1, w_2 \in T_p M$, definamos $\langle w_1, w_2 \rangle_p := \langle w_1, w_2 \rangle$.

- Definamos la forma cuadrática $I_p : T_p M \rightarrow \mathbb{R}$ dada por:

$$I_p(w) = \langle w, w \rangle_p = \langle w, w \rangle.$$

A I_p le llamaremos 1^{ra} Forma fundamental.

- Sea $(\mathbb{V}, +, \cdot)$ un espacio vectorial, una forma cuadrática sobre \mathbb{V} es $h : \mathbb{V} \rightarrow \mathbb{R}$ dada por:

$$h(v) = \sum_{i,j=1}^m a_{i,j} v_i v_j,$$

con $v = (v_1, v_2, \dots, v_m)$ y $h_{i,j}$ son escalares si $H = (a_{i,j})$.

1.5. La primera forma fundamental

Sea $S \subset \mathbb{R}^3$ una superficie regular, $p \in S$ y consideremos $T_p S \subset \mathbb{R}^3$. Como en \mathbb{R}^3 tenemos el producto interno $\langle \cdot, \cdot \rangle$ usual, entonces dado $v_p, w_p \in T_p S \subset \mathbb{R}^3$ podemos calcular:

$$\langle v_p, w_p \rangle = \sum_{i=1}^3 v_i w_i.$$

Así, al variar p en S podemos definir $\langle v_p, w_p \rangle_p = \langle v, w \rangle$.

Ahora bien, con $\langle \cdot, \cdot \rangle_p$ que es una forma bilineal, se define una forma cuadrática $I_p : T_p S \rightarrow \mathbb{R}$ dada por:

$$I_p(w) = \langle w_p, w_p \rangle_p = \|w\|^2 \geq 0.$$

A ésta le llamaremos primera forma fundamental de la superficie regular $S \subset \mathbb{R}^3$ en el punto $p \in S$.

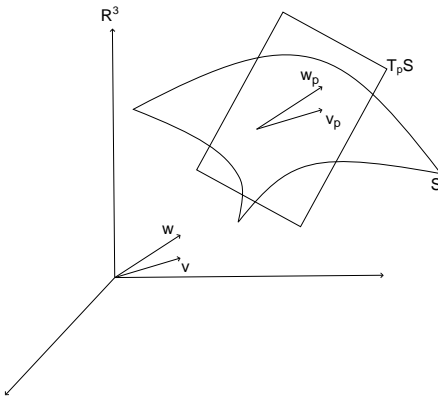


Figura 1.1: Métrica Riemanniana

Observemos que al cambiar p , cambia I_p y expresemos la primera forma fundamental en términos de la base $\{X_u, X_v\}$ asociada a la parametrización $X(u, v)$ en p . Como $w \in T_p S$, entonces $w = \alpha'(0)$ para alguna curva $\alpha : (-\epsilon, \epsilon) \rightarrow S$ con $\alpha(0) = p$; por otro lado, tenemos que:

$$\alpha(t) = X(u(t), v(t)).$$

Así, $p = \alpha(0) = X(u(0), v(0)) = X(u_0, v_0)$.

$$\begin{aligned}
 I_p(w) &= I_p(\alpha'(0)) \\
 &= \langle \alpha'(0), \alpha'(0) \rangle_p \\
 &= \langle u'(0)X_u(u_0, v_0) + v'(0)X_v(u_0, v_0), u'(0)X_u(u_0, v_0) + v'(0)X_v(u_0, v_0) \rangle_p \\
 &= \langle X_u(u_0, v_0), X_u(u_0, v_0) \rangle (u'(0))^2 + 2\langle X_u(u_0, v_0), X_v(u_0, v_0) \rangle u'(0)v'(0) \\
 &\quad + \langle X_v(u_0, v_0), X_v(u_0, v_0) \rangle (v'(0))^2 \\
 &= E(u'(0))^2 + 2Fu'(0)v'(0) + G(v'(0))^2 \\
 &= g_{11}(u'(0))^2 + 2g_{12}u'(0)v'(0) + g_{22}(v'(0))^2
 \end{aligned}$$

Donde,

$$\begin{aligned}
 g_{11} &= E = \langle X_u(u_0, v_0), X_u(u_0, v_0) \rangle_p, \\
 g_{12} &= g_{21} = F = \langle X_u(u_0, v_0), X_v(u_0, v_0) \rangle_p, \\
 g_{22} &= G = \langle X_v(u_0, v_0), X_v(u_0, v_0) \rangle_p.
 \end{aligned}$$

Más adelante la métrica Riemanniana será dada así: $\mathbf{G} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix}$

Ejemplo: Si $M \subset \mathbb{R}^3$ es un plano que pasa por $p_0 = (x_0, y_0, z_0)$ y con vectores directores $w_1 = (a_1, a_2, a_3)$ y $w_2 = (b_1, b_2, b_3)$ ortonormales, entonces:

$X(u, v) = p_0 + uw_1 + vw_2$, es parametrización de dicho plano. Calculemos G :

$$g_{11}(u, v) = \langle X_u, X_u \rangle = \langle w_1, w_1 \rangle = \|w_1\|^2 = 1, \forall u, v \in \mathbb{R}^2.$$

$$g_{12}(u, v) = g_{21}(u, v) = \langle X_u, X_v \rangle = \langle w_1, w_2 \rangle = 0.$$

$$g_{22}(v, v) = \langle X_v, X_v \rangle = \langle w_2, w_2 \rangle = \|w_2\|^2 = 1, \forall u, v \in \mathbb{R}^2.$$

Así,

$$G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Ejemplo: Sea $M = \{(x, y, z) \in \mathbb{R}^3 / x^2 + y^2 = 1\}$

$$X_1(u, v) = (\cos u, \sin u, v)$$

$$U_1 = \{(u, v) \in \mathbb{R}^2 / 0 < u < 2\pi, -\infty < v < \infty\}$$

$$X_2(u, v) = (\cos u, \sin u, v)$$

$$U_2 = \{(u, v) \in \mathbb{R}^2 / \pi < u < 3\pi, -\infty < v < \infty\}$$

Calculemos G :

$$g_{11}(u, v) = \langle X_u, X_u \rangle = \langle (-\sin u, \cos u, 0), (-\sin u, \cos u, 0) \rangle = \sin^2(u) + \cos^2(u) = 1, \forall (u, v) \in \mathbb{R}^2.$$

$$g_{12}(u, v) = g_{21}(u, v) = \langle X_u, X_v \rangle = \langle (-\sin u, \cos u, 0), (0, 0, 1) \rangle = 0.$$

$$g_{22}(v, v) = \langle X_v, X_u \rangle = \langle (0, 0, 1), (0, 0, 1) \rangle = 1, \forall (u, v) \in \mathbb{R}^2. \text{ Así,}$$

$$G = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Ejemplo: $S^2 = \{(x, y, z) \in \mathbb{R}^3 / x^2 + y^2 + z^2 = 1\}$. Sean,

$$X(\Theta, \rho) = (\sin \Theta \cos \rho, \sin \Theta \sin \rho, \sin \Theta)$$

$$X_\Theta(\Theta, \rho) = (\cos \Theta \cos \rho, \cos \Theta \sin \rho, -\sin \Theta)$$

$$X_\rho(\Theta, \rho) = (-\sin \Theta \sin \rho, \sin \Theta \cos \rho, 0)$$

Calculemos G :

$$g_{11}(\Theta, \rho) = \langle X_\Theta, X_\Theta \rangle$$

$$= \langle (\cos \Theta \cos \rho, \cos \Theta \sin \rho, -\sin \Theta), (\cos \Theta \cos \rho, \cos \Theta \sin \rho, -\sin \Theta) \rangle$$

$$= \cos^2 \Theta \cos^2 \rho + \cos^2 \Theta \sin^2 \rho + \sin^2 \Theta$$

$$= \cos^2 \Theta (\cos^2 \rho + \sin^2 \rho) + \sin^2 \Theta, \forall (\Theta, \rho) \in \mathbb{R}^2$$

$$\begin{aligned}
g_{12}(\Theta, \rho) &= g_{21}(\Theta, \rho) = \langle X_\Theta, X_\Theta \rangle \\
&= \langle (\cos \Theta \cos \rho, \cos \Theta \sin \rho, -\sin \Theta), (\cos \Theta \cos \rho, \cos \Theta \sin \rho, -\sin \Theta) \rangle \\
&= 0
\end{aligned}$$

$$\begin{aligned}
g_{22}(\Theta, \rho) &= \langle X_\rho, X_\rho \rangle \\
&= \langle (-\sin \Theta \sin \rho, \sin \Theta \cos \rho, 0), (-\sin \Theta \sin \rho, \sin \Theta \cos \rho, 0) \rangle \\
&= \sin^2 \Theta \sin^2 \rho + \sin^2 \Theta \cos^2 \rho = \sin^2 \Theta (\sin^2 \rho + \cos^2 \rho) \\
&= \sin^2 \Theta, \forall (\Theta, \rho) \in \mathbb{R}^2
\end{aligned}$$

Así,

$$G = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2 \Theta \end{pmatrix}$$

Definición 1.10. *(La Función de Gauss) Dado una parametrización $X : U \subset \mathbb{R}^2 \rightarrow M$ sobre una superficie M en el punto $p \in M$. Nosotros podemos escoger un vector normal unitario en cada punto de $X(U)$, dado por la siguiente regla:*

$$N(q) = \frac{X_u \times X_v}{\|X_u \times X_v\|}(q), \quad q \in X(U).$$

Así, se tiene una aplicación $N : X(U) \subset \mathbb{R}^3$ diferenciable, el cual asocia a cada $q \in X(U)$ un vector normal unitario $N(q)$. Queremos que tal aplicación sea diferenciable, para ello se necesita que la superficie sea orientable. A tal aplicación la llamaremos *orientación de M* cuando sea posible definirla globalmente.

Orientabilidad

Definición 1.11. *Un atlas de una superficie M , es un conjunto:*

$\mathbb{A} = \{\varphi : V_0 \rightarrow V, \text{ tal que } \varphi \text{ es parametrización de } V \text{ y la unión de tales } V \text{ cubre a } M\}$.

Definición 1.12. Dos parametrizaciones $\varphi : V_0 \rightarrow V, \psi : W_0 \rightarrow W$, se dicen compatibles cuando $V \cap W \neq \emptyset$ y $\psi \circ \varphi : \varphi^{-1}(V \cap W) \rightarrow \psi^{-1}(V \cap W)$ tiene $\det J(\psi^{-1} \circ \varphi)(x) > 0, \forall x \in \varphi^{-1}(V \cap W)$.

Definición 1.13. Un atlas se dice coherente cuando, para todo par de parametrizaciones son compatibles.

Definición 1.14. Una superficie se dice orientable cuando admite un atlas coherente.

Teorema 1.3. Sea $M \subset \mathbb{R}^{m+n}$ una superficie de co-dimensión n (es decir, $\dim M = m$) y admite n campos vectoriales continuos Linealmente independientes de vectores normales $v_1, v_2, \dots, v_n : M \rightarrow \mathbb{R}^{m+1}$ si y solo si M es orientable.

Proposición 1.2. Si $M = f^{-1}(c)$ es la imagen inversa de un valor regular de $f : U \subset \mathbb{R}^{m+n} \rightarrow \mathbb{R}^n$ de clase C^k , entonces M es orientable.

1.6. Geometría de la Aplicación de Gauss

Definición 1.15. Sea $M \subset \mathbb{R}^3$ es una superficie con orientación N . Consideremos la función $N : M \rightarrow \mathbb{R}^3$ toma sus valores en la esfera unitaria:

$$S = \{(x, y, z) \in \mathbb{R}^3 / x^2 + y^2 + z^2 = 1\},$$

a tal aplicación se le llama Aplicación de Gauss de S .

Si $M \subset \mathbb{R}^3$ es orientable y N es una orientación consideremos:

$dN_p : T_p(M) \rightarrow T_{N(p)}S^2$ pero $T_p(M) \parallel T_{N(p)}S^2$, es por ello que se acostumbra a escribir:

$$dN_p : T_p M \rightarrow T_p M$$

Ejemplo:

Para el plano $ax + by + cz = 0$, el campo normal viene dado por:

$$N(p) = \frac{(a, b, c)}{\|(a, b, c)\|} = \frac{(a, b, c)}{\sqrt{a^2 + b^2 + c^2}}, \forall p \text{ en el plano.}$$

Ejemplo:

Para $M^2 = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$.

Sea $\alpha : (-\epsilon, \epsilon) \rightarrow S^2$ una curva diferenciable en M^2 parametrizada por $\alpha(t) = (x(t), y(t), z(t))$ tal que:

$$x^2(t) + y^2(t) + z^2(t) = 1.$$

Derivando tenemos:

$$2x(t)x'(t) + 2y(t)y'(t) + 2z(t)z'(t) = 0.$$

Esto es,

$$\langle \alpha(t), \alpha'(t) \rangle' = 0.$$

Luego,

$$N(\alpha(t)) = \alpha(t).$$

Derivando,

$$N'(\alpha(t))\alpha'(t) = \alpha'(t) \Rightarrow dN_{\alpha(t)}\alpha'(t) = \alpha'(t).$$

Por tanto,

$$dN_p(v) = v. \text{ (Identidad)}$$

Definición 1.16. Llamamos 2^{da} forma fundamental a la forma:

$$II_p(v) = -\langle dN_p(v), v \rangle.$$

Definición 1.17. Sea $p \in M$ y $dN_p : T_p(M) \rightarrow T_p(M)$, el determinante $\det[dN_p]$, se le llama curvatura Gaussiana, la cual denotaremos por k .

Definición 1.18. Un difeomorfismo $\varphi : M \rightarrow \overline{M}$ es un isometría si y solo si para todo $p \in M$ y $\forall w_1, w_2 \in T_p(M)$, se cumple que:

$$\langle w_1, w_2 \rangle_p = \langle d\varphi_p(w_1), d\varphi_p(w_2) \rangle_{\varphi(p)}.$$

Así, diremos que \overline{M} y M son isométricas. De aquí que:

$$I_p(w) = \langle w, w \rangle = \langle d\varphi_p(w), d\varphi_p(w) \rangle = I_{\varphi_p}(d\varphi_p(w)).$$

Definición 1.19. Una $\varphi : V \subset M \rightarrow \overline{M}$, con V vecindad de $p \in M$ es isometría local si y solo si $\exists \overline{V} \subset \overline{M}$ tal que $\varphi : V \rightarrow \overline{V}$ es isometría.

Definición 1.20. Un difeomorfismo $\varphi : M \rightarrow \overline{M}$ es llamada una aplicación conforme si $\forall p \in M$ y $\forall v_1, v_2 \in T_p(M)$, se tiene que:

$$\langle d\varphi_p(v_1), d\varphi_p(v_2) \rangle_{\varphi(p)} = \lambda^2(p) \langle v_1, v_2 \rangle_p, \text{ donde } \lambda^2 \text{ es no-nula y diferenciable.}$$

Definición 1.21. Un campo en un conjunto abierto $U \subset \mathbb{R}^2$ es una aplicación que asigna a cada $q \in U$ un vector $w(q) \in \mathbb{R}^2$. Diremos que W es un campo vectorial diferenciable si para cada $q = (x, y) \in U$ podemos escribir:

$$W(q) = \left(a(x, y), b(x, y) \right) \text{ y las funciones } a \text{ y } b \text{ son diferenciables en } U.$$

Definición 1.22. Sea $X : U \subset \mathbb{R}^2 \rightarrow M$ una parametrización en la orientación de M . Es posible asignar a cada punto de $X(U)$ un triedro natural dado por los vectores X_u, X_v y N . Expresemos la derivada de los vectores X_u, X_v y N en la base $\{X_u, X_v, N\}$.

$$X_{uu} = \Gamma_{11}^1 X_u + \Gamma_{11}^2 X_v + L_1 N.$$

$$X_{uv} = \Gamma_{12}^1 X_u + \Gamma_{12}^2 X_v + L_2 N.$$

$$X_{vu} = \Gamma_{21}^1 X_u + \Gamma_{21}^2 X_v + \bar{L}_2 N.$$

$$X_{vv} = \Gamma_{22}^1 X_u + \Gamma_{22}^2 X_v + L_3 N$$

$$N_u = a_{11} X_u + a_{21} X_v$$

$$N_v = a_{12} X_u + a_{22} X_v.$$

Los coeficientes Γ_{ij}^k se les llaman símbolos de Cristoffel de S en la parametrización X .

Ya que $X_{uv} = X_{vu}$, concluimos que $\Gamma_{12}^1 = \Gamma_{21}^1$ y $\Gamma_{12}^2 = \Gamma_{21}^2$; por otro lado, si hacemos el producto interno entre N y X_{uu} obtenemos $\langle N, X_{uu} \rangle = L_1 \langle N, N \rangle$.

Así, $L_1 = \frac{\langle N, X_{uu} \rangle}{\langle N, N \rangle} = e$.

Haciendo el producto interno con las primeras cuatro ecuaciones anteriores y N , obtenemos que $L_1 = e$, $L_2 = \bar{L}_2 = f$ y $L_3 = G$, donde e , f y g son los coeficientes de la segunda forma fundamental.

1.7. Superficies Abstractas y Variedades Riemannianas.

Definición 1.23. Una Variedad diferenciable de dimension n es un conjunto M junto con una familia de funciones uno a uno, $X_\alpha : U_\alpha \rightarrow M$, donde $U_\alpha \subset (\mathbb{R})$ son abiertos en M tales que:

1. $\bigcup_\alpha X_\alpha(U_\alpha) = M$.
2. $\forall \alpha, \beta$ con $X_\alpha(U_\alpha) \cap X_\beta(U_\beta) = M \neq \emptyset$ se tiene que: $X_\alpha^{-1} \circ X_\beta$ y $X_\beta^{-1} \circ X_\alpha$ son diferenciables.

3. La familia $\{U_\alpha, X_\alpha\}$ (estructura diferenciable para M) es maximal relativo a 1 y 2.

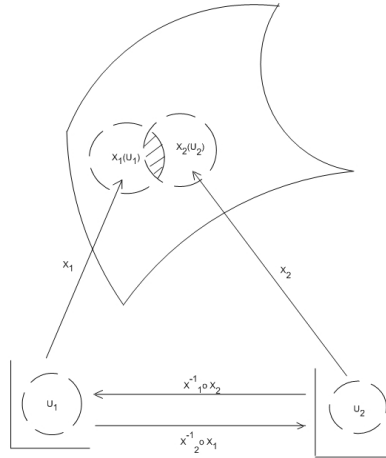


Figura 1.2: Variedad Diferencial.

Definición 1.24. Una función diferenciable $\alpha : (-\epsilon, \epsilon) \rightarrow M$ es llamada curva sobre M . Supongamos que $\alpha(0) = p$, $D = \{f : M \rightarrow \mathbb{R} / f \text{ es diferenciable en } p\}$. El vector tangente a α en $t = 0$ es la función:

$$\alpha'(0) : D \rightarrow \mathbb{R}$$

dada por:

$$\begin{aligned} \alpha'(0)(f) &= \left. \frac{d}{dt}(f \circ \alpha) \right|_{t=0}, f \in D \\ &= u'(0) \left(\frac{\partial f}{\partial u} \right)_0 + v'(0) \left(\frac{\partial f}{\partial v} \right)_0 \\ &= \left\{ u'(0) \left(\frac{\partial}{\partial u} \right)_0 + v'(0) \left(\frac{\partial}{\partial v} \right)_0 \right\} f. (\text{Notacion}) \end{aligned}$$

Definición 1.25. El conjunto de vectores tangentes a M en p sera indicado por $T_p M$. Este conjunto $T_p M$ con las operaciones usuales de funciones forman

un espacio vectorial de dimensión n y, escogiendo una parametrización, $X : U \rightarrow M$ se determina una base asociada a ésta, la cual se denota por:

$$\left\{ \left(\frac{\partial}{\partial x} \right)_o, \dots, \left(\frac{\partial}{\partial x_n} \right)_o \right\}.$$

y es posible verificar que una estructura lineal en T_pM , así definida no depende de la parametrización. Este espacio vectorial es llamado espacio tangente a M en p .

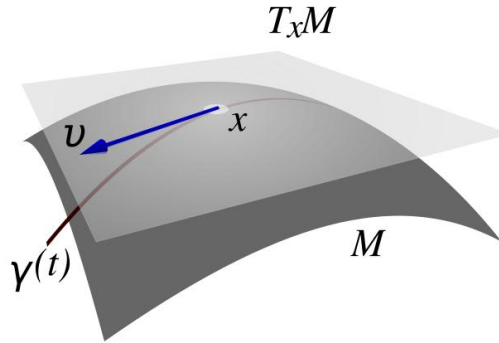


Figura 1.3: Plano Tangente.

Definición 1.26. Sean M_1, M_2 variedades y sea $\varphi : M_1 \rightarrow M_2$ diferenciable $\forall p \in M_1, \forall w \in T_p(M_1)$ consideremos $\alpha : (-\epsilon, \epsilon) \rightarrow M_1$ con $\alpha(0) = p$ y $\alpha'(0) = w$. Sea $\beta = \varphi \circ \alpha$ y la función: $d\varphi_p : T_p(M_1) \rightarrow T_{\varphi(p)}(M_2)$ dada por $d\varphi_p(w) = \beta'(0)$, esta bien definida y es lineal, la cual llamaremos la diferencial de φ en p .

Definición 1.27. Sea M una variedad diferenciable de dimensión n y sea $TM : \{(p, v); p \in M, v \in T_pM\}$. Si dotamos a este conjunto de una estructura diferenciable, este sera una variedad de dimensión $2n$, dicho conjunto con tal estructura es llamado **Fibrado Tangente** denotado por:

$$TM = \bigcup_{p \in M} T_pM.$$

Definición 1.28. Una métrica Riemanniana (o estructura Riemanniana) en una variedad diferencial M es una correspondencia que asocia a cada punto $p \in M$ un producto interno $\langle \cdot, \cdot \rangle_p$, en el espacio tangente $T_p M$ que varia diferencialmente en el siguiente sentido, sea $X : U \subset \mathbb{R}^n \rightarrow M$ un sistema de coordenadas locales en torno a p , con $x(x_1, x_2, \dots, x_n) = q \in x(U)$ y $\frac{\partial}{\partial x_i}(q) = d_x(0, \dots, 1, \dots, 0)$ entonces,

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

es una función diferenciable en U .

Definición 1.29. El producto interno de dos vectores $u, v \in T_p M$ se escribe:

$$\langle u, v \rangle_p := g_p(u, v),$$

donde g_p es la métrica en el punto p .

Definición 1.30. La norma de un vector $v \in T_p M$, es definida por:

$$\|v\| := \sqrt{\langle v, v \rangle_p}.$$

Definición 1.31. Una superficie Geométrica o Variedad Riemanniana de dimensión α , M , es una variedad junto a la escogencia de un producto interno $\langle \cdot, \cdot \rangle_p$ en cada $T_p(M)$, $p \in M$ el cual varia diferencialmente con p en el siguiente sentido, para alguna (y por tanto para todas) parametrización $X_\alpha : U_\alpha \rightarrow M$ las funciones:

- $E(u, v) = g_{11} = \langle \frac{\partial}{\partial u}, \frac{\partial}{\partial u} \rangle.$
- $F(u, v) = g_{12} = g_{21} = \langle \frac{\partial}{\partial u}, \frac{\partial}{\partial v} \rangle.$
- $G(u, v) = g_{22} = \langle \frac{\partial}{\partial v}, \frac{\partial}{\partial v} \rangle.$

son funciones diferenciables en U_α . Al producto interno $\langle \cdot, \cdot \rangle$ se le llama métrica Riemanniana sobre M .

Definición 1.32. Una función diferenciable $\varphi: M \rightarrow \mathbb{R}$ de una variedad, es una inmersión si $d\varphi_p : T_p(M) \rightarrow T_p(\mathbb{R})$ es inyectiva y si además M tiene métrica \langle, \rangle y,

$$\langle d\varphi_p(v), d\varphi_p(w) \rangle_{d\varphi_p} = \langle v, w \rangle_p,$$

se dice que φ es una inmersión isométrica. Si φ es una inmersión y es homeomorfismo sobre su imagen decimos que φ es un embedding (encaje).

Definición 1.33. $V \subset M$ es abierto en M si y solo si $\forall \alpha : X_\alpha^{-1}(V \cap X_\alpha(U_\alpha))$ es abierto en \mathbb{R} .

Definición 1.34. $f : M \rightarrow \mathbb{R}$ sera diferenciable si y solo si $\forall \alpha$ tal que $f \circ X_\alpha : U_\alpha \subset \mathbb{R} \rightarrow \mathbb{R}$ es diferenciable.

Definición 1.35. Sea $D = \{f : M \rightarrow \mathbb{R} : f \text{ es diferenciable} \}$ y sea $\alpha : (-\epsilon, \epsilon) \rightarrow M$ curva diferenciable. Llamaremos vector tangente a α en $t = 0$ a la función $\alpha'(o) : D \rightarrow \mathbb{R}$, dada por: $\alpha'(o)(f) = \left. \frac{d}{dt}(f \circ \alpha) \right|_{t=0}$. Una base para $T_p(M) = \{\alpha'(0)/\alpha : (-\epsilon, \epsilon) \rightarrow M \text{ es curva diferenciable}\}$ seria:

$$\left\{ \left(\frac{\partial}{\partial x_1} \right)_o, \left(\frac{\partial}{\partial x_2} \right)_o, \dots, \left(\frac{\partial}{\partial x_n} \right)_o \right\}.$$

Definición 1.36. Una variedad Riemanniana es una variedad diferenciable n -dimensional M junto a una escogencia, $\forall p \in M$, de un producto interno \langle, \rangle_p en $T_p M$ que varia diferencialmente con el punto p , esto es: para alguna parametrización (por tanto para toda) $X_\alpha : U_\alpha \rightarrow M$ con $p \in X_\alpha(U_\alpha) \subset M$ las funciones,

$$g_{ij}(x_1, \dots, x_n) = \left\langle \left(\frac{\partial}{\partial x_i} \right)_o, \left(\frac{\partial}{\partial x_j} \right)_o \right\rangle; i, j = 1, 2, \dots, n$$

son diferenciables en $X'_\alpha(p)$ y x_1, x_2, \dots, x_n son coordenadas de $U_\alpha \subset \mathbb{R}$.

La familia $\{\langle, \rangle_p : p \in M\}$ la llamaremos estructura Riemanniana o métrica Riemanniana para M . Notación: $G = (g_{ij})$.

Si M es una variedad diferencial dotada de una métrica Riemanniana g , entonces M es una variedad Riemanniana y la denotamos por (M, g) , o solo por M si no surge confusión donde G denota la matriz de representación de la métrica g . Cualquier variedad (M, G) puede ser convertida en un espacio métrico (M, d) donde d es la métrica inducida por la métrica Riemanniana g .

Definición 1.37. La longitud de una sección curva (trozo) de una curva plana $\alpha : [t_0, t_1] \rightarrow M$ que une $\alpha(t_0) = p'$ con $\alpha(t_1) = p$, es definida mediante la métrica como:

$$L(\alpha) = \int_{t_0}^{t_1} \|\alpha'(t)\| dt.$$

Definición 1.38. La distancia Riemanniana entre dos puntos p' y p en M se define como:

$$d(p', p) = \inf\{L(\alpha) : \alpha \text{ es una curva diferenciable que conecta a } p' \text{ con } p\}.$$

1.8. Transporte Paralelo y Geodésica

Definición 1.39. Sea $\alpha : I \rightarrow M$ una curva parametrizada en M . Un campo vectorial W a lo largo de α es una correspondencia que asigna a cada punto $t \in I$ un vector $W(t) \in T_{\alpha(t)}(M)$. Diremos que W es un campo diferenciable en $t \in I$ si y sólo si existe $\varphi(u, v)$ parametrización en $\alpha(t_0)$ tal que:

$W(t) = a\varphi_u + b\varphi_v$ en la base $\{\varphi_u, \varphi_v\}$, y las aplicaciones $a(t)$ y $b(t)$ son diferenciables.

Definición 1.40. Sea W un campo vectorial diferenciable sobre el conjunto abierto $U \subset M$ un abierto de la superficie M y $p \in M$. Sea $y \in T_p M$, consideremos una curva parametrizada $\alpha : (-\epsilon, \epsilon) \rightarrow U$, con $\alpha(0) = p$ y $\alpha'(0) = y$. Sea $W(t)$ la restricción de W a la curva $\alpha(t)$. Llamaremos derivada covariante del campo W respecto al vector y en el punto p al vector obtenido al proyectar $(\frac{dW}{dt})(0)$ normalmente sobre el plano $T_p M$. La derivada covariante es denotada por $(\frac{DW}{dt})(0)$ o $(D_y W)(p)$.

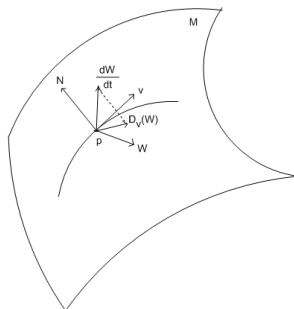


Figura 1.4: Derivada Covariante

Definición 1.41. Una curva parametrizada definida en $[0, \ell]$, $\alpha : [0, \ell] \rightarrow M$ es la restricción a $[0, \ell]$ de una función diferenciable $\alpha : (0 - \epsilon, \ell + \epsilon) \rightarrow M$, $\epsilon > 0$. Si $\alpha(0) = p$ y $\alpha(\ell) = q$ decimos que α conecta p con q .

- α es regular sí y solo si $\alpha'(t) \neq 0, \forall t \in [0, \ell]$.

Definición 1.42. Sea W un campo vectorial diferenciable a lo largo de $\alpha : I \rightarrow M$, la expresión $\frac{DW}{dt}(t)$, $t \in I$ está bien definida y es llamada derivada covariante de W en t .

Definición 1.43. Un campo vectorial W a lo largo de una curva parametrizada $\alpha : I \rightarrow M$ es paralelo si $\frac{DW}{dt}(t) = 0$ para todo $t \in I$. Esto es, cuando su proyección a lo largo del normal es un punto ó si todos los vectores son paralelos a los vectores normales.

Proposición 1.3. Sea V y W campos vectoriales paralelos a lo largo de $\alpha : I \rightarrow M$, entonces $\langle W(t), V(t) \rangle$ es constante. En particular, $\|W(t)\| = ctte$, $\|V(t)\| = ctte$ y $\angle(W(t), V(t))$ es constante.

Demostración:

Sean V y W campos vectoriales paralelos a lo largo de una curva $\alpha : I \rightarrow M$, entonces: $\frac{DW}{dt} = 0 = \frac{DV}{dt}$ y $W'(t) = \frac{dW(t)}{dt} = \lambda_1 N$ y $V'(t) = \frac{dV(t)}{dt} = \lambda_2 N$.

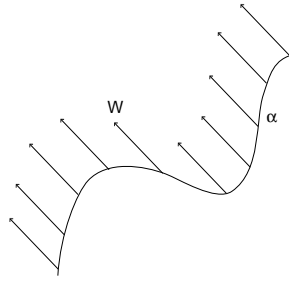


Figura 1.5: Campo Vectorial Paralelo

Así;

$$\langle W'(t), V(t) \rangle = 0 \text{ y } \langle W(t), V'(t) \rangle = 0.$$

Derivando $\langle W(t), V(t) \rangle$ nos queda:

$$(\langle W(t), V(t) \rangle)' = \langle W'(t), V(t) \rangle + \langle W(t), V'(t) \rangle = 0.$$

Por lo tanto, $\langle W(t), V(t) \rangle = \text{ctte.}$ □

Proposición 1.4. *Sea $\alpha : I \rightarrow M$ una curva parametrizada en M y sea $w_0 \in T_{\alpha(t_0)}M$, $t_0 \in I$ entonces existe un único campo vectorial $W(t)$ a lo largo de $\alpha(t)$, con $W(t_0) = W_0$. Al vector W_1 se llama transporte paralelo de $W(t_0)$ a lo largo de α en t_1 .*

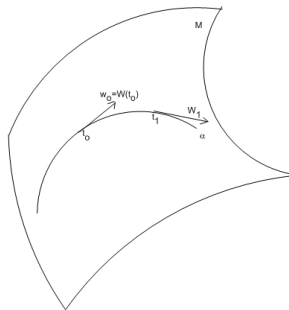


Figura 1.6: Transporte Paralelo

Definición 1.44. Dado dos campos vectoriales V y W en M , la derivada covariante de W en la dirección de V se denota por $\nabla_v W$.

En éste trabajo ∇ es la conexión Levi-Civita asociado a (M, G) . Esta conexión define una única derivada covariante $\frac{D}{dt}$, donde para cada campo vectorial V , a lo largo de una curva $\alpha : [t_0, t_1] \rightarrow M$, otro campo vectorial se obtiene, denotado por $\frac{DV}{dt}$.

Definición 1.45. Una conexión afín ∇ en una variedad diferenciable M es una aplicación:

$$\nabla : X(M) \times X(M) \rightarrow X(M)$$

que se indica por: $(X, Y) \rightarrow \nabla_X Y$ y que satisface las siguientes propiedades:

- $\nabla_{fX+gY} Z = f\nabla_X Z + g\nabla_Y Z$
- $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$
- $\nabla_X (fY) = f\nabla_X Y + X(f)Y$

donde $X, Y, Z \in X(M)$ (conjunto de campos de vectores de clase C^∞ en M) y $f, g \in D(M)$ (conjunto de funciones de clase C^∞ definidas en M).

Definición 1.46. Sea M una variedad diferenciable con una conexión afín ∇ . Sea $\alpha : I \rightarrow M$ una curva diferenciable en M y v_o un vector tangente a M en $\alpha(t_0), t_o \in I$ (es decir, $v_o \in T_{\alpha(t_0)}M$), al único campo de vectores paralelo X a lo largo de α tal que $X(t_0) = v_o$, es llamado transporte paralelo de v_o a lo largo de α .

En este trabajo, el transporte paralelo a lo largo de α , desde $\alpha(t_0)$ hasta $\alpha(t_1)$, lo denotaremos por P_{α, t_0, t_1} , tal que:

$$P_{\alpha, t_0, t_1} : T_{\alpha(t_0)}M \rightarrow T_{\alpha(t_1)}M, \text{ definido por } P_{\alpha, t_0, t_1}(v) = v(t_1),$$

donde V es el único campo vectorial a lo largo de α tal que $\frac{DV}{dt} = 0$ y $V(t_0) = v$. Además,

$$P_{\alpha,t_0,t_1}^{-1} = P_{\alpha,t_1,t_0} \text{ y } P_{\alpha,t_0,t_1} = P_{\alpha,t,t_1} P_{\alpha,t_0,t}, \forall t \in [t_0, t_1].$$

Definición 1.47. Fijemos $p, q \in S$ y $\alpha : I \rightarrow S$ con $\alpha(0) = p$ y $\alpha(1) = q$. Denotemos por $\tau_\alpha(v) : T_p(S) \rightarrow T_q(S)$ la función que asigna a cada $v \in T_p(S)$ un transporte paralelo a lo largo de α en q .

Definición 1.48. Una curva parametrizada, no constante $\gamma : I \rightarrow S$ se dice geodésica en $t \in I$ si el campo vectorial tangente $\gamma'(t)$ es paralelo a lo largo de γ en t ; esto es que $\frac{D\gamma'(t)}{dt} = 0$.

γ es geodésica si y sólo si γ es geodésica en todo $t \in I$.

Si $[a, b] \subset I$ y $\gamma : I \rightarrow M$ es geodésica, la restricción de γ a $[a, b]$ se llama el segmento geodésico uniendo $\gamma(a)$ con $\gamma(b)$.

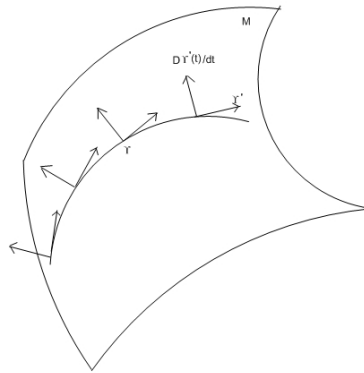
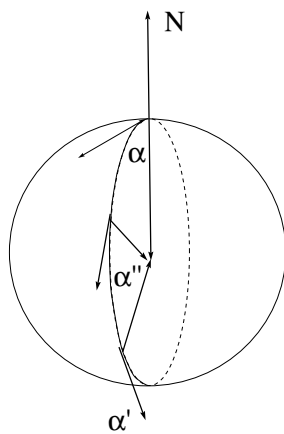


Figura 1.7: Geodésica

Definición 1.49. Un segmento geodésico $\gamma : [a, b] \rightarrow M$ se llama minimizante si $l(\gamma) \leq l(c)$, donde $l(\cdot)$ indica la longitud de una curva y c es cualquier curva diferenciable por partes uniendo a $\gamma(a)$ con $\gamma(b)$.

Ejemplo:

Supongamos que S es una superficie regular y que $\alpha : I \rightarrow M$ es una curva diferenciable en M , donde $\alpha(0) = p$ y $\alpha'(0) = \gamma$. Consideremos el campo W formado por todos $\alpha'(0) = \gamma \in T_p M$ a lo largo de la curva; así $\frac{dW}{dt} = \alpha''(0)$. Si consideramos el caso particular en el cual $\alpha''(0) = \frac{dW}{dt}$ es paralelo al normal N a lo largo de la curva α , entonces $\frac{DW}{dt}(t) = 0$, para cualquier $t \in I$.



Lema 1.1. Si $\gamma(t, \mathbf{v})$ es geodésica, $t \in (-\epsilon, \epsilon)$ entonces la geodésica $\gamma(t, \lambda)$, $\lambda \in \mathbb{R}$, $\lambda \neq 0$, está definida para $t \in (\frac{-\epsilon}{\lambda}, \frac{\epsilon}{\lambda})$ y $\gamma(t, \lambda \mathbf{v}) = \gamma(\lambda t, \mathbf{v})$.

Demostación:

Sea $\alpha : (\frac{-\epsilon}{\lambda}, \frac{\epsilon}{\lambda}) \rightarrow M$ una curva parametrizada definida por $\alpha(t) = \gamma(\lambda t)$, entonces $\alpha'(0) = \lambda \gamma'(0) = \lambda \mathbf{v}$. Por la linealidad de la derivada covariante se tiene que: $D_{\alpha'(t)} \alpha'(t) = D_{\lambda \gamma'(t)} \lambda \gamma'(t) = \lambda^2 D_{\gamma'(t)} \gamma'(t)$. De acá se tiene que α es una geodésica con las siguientes condiciones iniciales $\gamma(0)$, $\lambda \gamma'(0)$. Por unicidad, $\alpha(t) = \gamma(t, \lambda \mathbf{v}) = \gamma(\lambda t, \mathbf{v})$. □

Notación: si $\mathbf{v} \in T_p S$ y $\mathbf{v} \neq \emptyset$ es tal que $\gamma(\|v\|, \frac{v}{\|v\|}) = \gamma(1, v)$, se define la exponencial dada como sigue: $exp_p(\mathbf{v}) = \gamma(1, \mathbf{v})$ y $exp_p(0) = p$.

Definición 1.50. Un triángulo geodésico $\Delta(x_1, x_2, x_3)$ en una variedad Riemannina es el conjunto que consiste de tres puntos distintos x_1, x_2, x_3 los

cuales son llamados los vertices y tres segmentos geodésicos minimales γ_{i+1} que une x_{i+1} con x_{i+2} los cuales son llamados lados, donde $i = 1, 2, 3$.

Definición 1.51. Una curva geodésica $\gamma : I \rightarrow M$ a partir del punto x en la dirección $v \in T_x M$, ($\gamma(0) = x, \gamma'(0) = v$) viene dada por:

$$\frac{d^2 \gamma_k}{dt^2} + \sum_{i,j=1}^n \gamma_{ij}^k \frac{d\gamma_i}{dt} \frac{d\gamma_j}{dt} = 0, k = 1, \dots, n$$

donde γ_{ij}^k son los simbolos de Christoffel's expresados por:

$$\gamma_{ij}^k = \frac{1}{2} \sum_{k=1}^n \left\{ \frac{\partial g_{jk}}{\partial x_i} + \frac{\partial g_{ki}}{\partial x_j} + \frac{\partial g_{ij}}{\partial x_k} \right\} g^{km},$$

donde (g^{ij}) denota la matriz inversa de la métrica $g = (g_{ij})$, y x_i es la coordenada de x .

Definición 1.52. Una variedad Riemanniana es completa si sus geodésicas se definen para cualquier valor $t \in \mathbb{R}$.

1.9. La Función Exponencial.

Definición 1.53. Dado $x \in M$ y $v \in T_x M, v \neq 0$, sabemos que existe una única geodésica $\gamma : (-\epsilon, \epsilon) \rightarrow M$ tal que $\gamma(0) = x$ y $\gamma'(0) = v$, entonces se define la exponencial, $\exp_x : T_x M \rightarrow M$, como sigue:

$$\exp_x(v) = \gamma(1, v) = \gamma(1) \text{ y } \exp_x(0) = x$$

Si M es completo, entonces \exp_x se define para todo $v \in T_x M$.

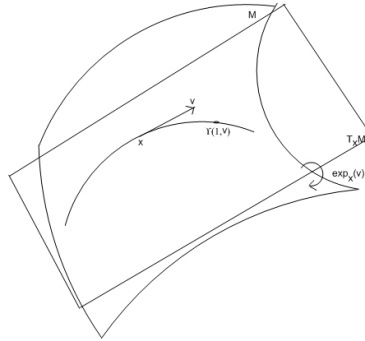


Figura 1.8: Función Exponencial.

Teorema 1.4. (Hopf-Rinow) Si M es completo, entonces para cada $p, q \in M$ existe una geodésica minimal, que une a p y q tal que su longitud es igual a la distancia de p a q , esto es: $\ell(\gamma) = d(p, q)$, donde:

$$d(p, q) = \inf\{\ell(\alpha) : \alpha \text{ es una curva que une a } p \text{ con } q\}.$$

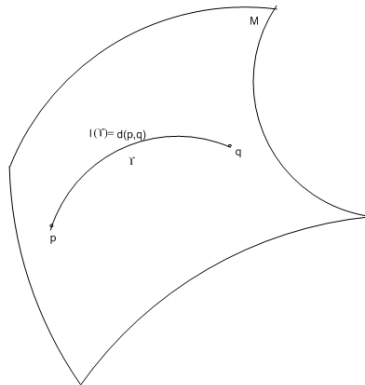


Figura 1.9: Teorema de Hopf-Rinow.

Demostración: ver [2]

Proposición 1.5. La exponencial, $exp_p : B_\epsilon \subset T_p(M) \rightarrow M$, es un difeomorfismo en una vecindad $U \subset B_\epsilon$ del origen de T_pM .

Demostración:

Consideremos la curva $\alpha : (-\epsilon, \epsilon) \rightarrow M$ tal que $\alpha(t) = tv$, $v \in T_p(M)$, $\alpha(0) = 0$, $\alpha'(t) = v$, $\alpha'(0) = v$, $\forall t \in (-\epsilon, \epsilon)$.

Luego,

$$(\exp_p \circ \alpha)(t) = \exp_p(\alpha(t)) = \exp_p(tv).$$

Derivando,

$$\left. \frac{d}{dt}(\exp_p(tv)) \right|_{t=0} = \left. \frac{d}{dt}(\gamma(t; v)) \right|_{t=0} = v \quad \square$$

Definición 1.54. Un tensor T de orden r en una variedad Riemanniana es una aplicación multilineal,

$$T : X(M) \times \dots \times X(M) \rightarrow D(M)$$

Esto quiere decir que, dados $Y_1, \dots, Y_r \in X(M)$, funciones diferenciales en M donde T es lineal en cada argumento, esto es,

$$T(Y_1, \dots, fX + gY, \dots, Y_r) = fT(Y_1, \dots, X, \dots, Y_r) + gT(Y_1, \dots, Y, \dots, Y_r),$$

para todo $X, Y \in X(M)$, $f, g \in D(M)$.

Definición 1.55. Dados los campos vectoriales $X, Y, Z \in M$, denotamos por R el campo tensor curvatura a la aplicación:

$$R : T_pM \times T_pM \rightarrow \mathbb{R}.$$

Definida por: $R(X, Y)Z = \nabla_y \nabla_x Z - \nabla_x \nabla_y Z + \nabla_{[X, Y]}Z$, donde $[X, Y]$ es el corchete de Lie.

Definición 1.56. Sean X e Y campos de vectores. La curvatura seccional con respecto a X e Y es definida por:

$$K(X, Y) = \frac{\langle R(X, Y)X, Y \rangle}{\|X\|^2 \|Y\|^2 - \langle X, Y \rangle^2},$$

donde, $\|X\| = \langle X, X \rangle^{\frac{1}{2}}$.

Si $K(X, Y) \geq 0$, entonces M es una variedad Riemanniana con curvatura seccional no-negativa.

Definición 1.57. El gradiente de una función, $f : M \rightarrow \mathbb{R}$, $\text{grad}f$, es un campo vectorial en M , se define por:

$$df(X) = \langle \text{grad}f, X \rangle,$$

donde X es también un campo vectorial en M .

Definición 1.58. Una bisagra geodésica $(p; \gamma_1, \gamma_2)$ en M es una figura que consiste en un punto $p \in M$ llamado vertice y segmentos geodésicos minimales γ_1 y γ_2 emanados de p llamados lados. Nosotros denotaremos por α el ángulo entre los vectores tangentes a γ_1 y γ_2 en p , el cual es llamado el ángulo de la bisagra geodésica $(p; \gamma_1, \gamma_2)$, si γ_1 es minimal, pero γ_2 no necesariamente minimal, llamaremos a está bisagra geodésica generalizada. Si $\ell_1 = \ell(\gamma_1)$, $\ell_2 = \ell(\gamma_2)$, $\ell_3 = \ell(\gamma_3)$ y $\alpha = \angle(\gamma'_1(0), \gamma'_2(0))$, obtenemos el siguiente resultado:

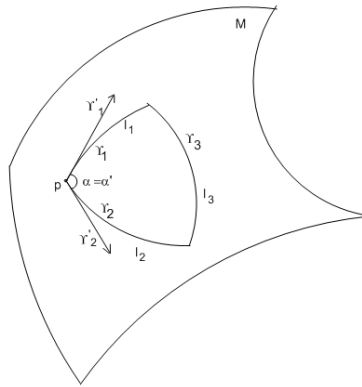


Figura 1.10: Bisagra Geodesica.

Teorema 1.5 (Toponogov). Sea M una variedad Riemanniana completa cuya curvatura seccional satisfaciendo que $K \geq 0$. Y consideremos T_pM variedad de curvatura constante 0. Sea $(p; \gamma'_1, \gamma'_2)$ una bisagra geodésica en T_pM tal que: $\ell(\gamma'_1) = \ell(\gamma_1)$, $\ell(\gamma'_2) = \ell(\gamma_2)$ y su ángulo α' es igual al de $(p; \gamma_1, \gamma_2)$. Sean q, s los puntos finales de γ_1 y γ'_1 respectivamente y sean r, t los puntos finales de γ_2 y γ'_2 respectivamente, entonces,

$$d(q, r) \leq d(s, t).$$

Demostración: ver [2]

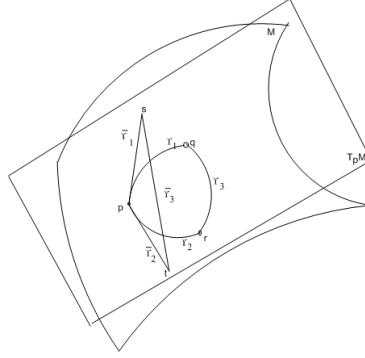


Figura 1.11: Teorema de Toponogov.

Teorema 1.6. *Sea M una variedad Riemanniana completa de dimensión finita, con curvatura seccional no-negativa tenemos que:*

$$\ell_3^2 \leq \ell_1^2 + \ell_2^2 - 2\ell_1\ell_2\cos\alpha,$$

donde ℓ_i denota la longitud de γ_i , ($i = 1, 2$), $\ell_3 = d(\gamma_1(\ell_1), \gamma_2(\ell_2))$ y $\alpha = \angle(\gamma_1'(0), \gamma_2'(0))$.

Demostración:

Dada una bisagra geodésica $(p; \gamma_1, \gamma_2)$ en M , se considerará su correspondiente geodésica en el plano $(\bar{p}; \bar{\gamma}_1, \bar{\gamma}_2)$ tal que $\bar{\ell}_i = \ell_i$, $i = 1, 2$, $\bar{\alpha} = \alpha$.

Debido a que las geodésicas en el plano son líneas rectas, podemos aplicar la ley usual de los cosenos a esta bisagra plana, así tenemos que:

$$\begin{aligned} \bar{\ell}_3^2 &= \bar{\ell}_1^2 + \bar{\ell}_2^2 - 2\bar{\ell}_1\bar{\ell}_2\cos\bar{\alpha} \\ &= \ell_1^2 + \ell_2^2 - 2\ell_1\ell_2\cos\alpha \quad (\text{por hipótesis}) \end{aligned}$$

Luego, por el teorema de Toponogov tenemos que $\ell_3 \leq \bar{\ell}_3$ y por lo tanto,

$$\ell_3^2 \leq \ell_1^2 + \ell_2^2 - 2\ell_1\ell_2\cos\alpha$$

□

Definición 1.59. Decimos que $f : M \rightarrow \mathbb{R}$ es convexa si para cada geodésica $\gamma : \mathbb{R} \rightarrow M$, $f \circ \gamma : \mathbb{R} \rightarrow \mathbb{R}$ es una función real convexa si se cumple:

$$f(\gamma((1 - \lambda)a + \lambda b)) \leq (1 - \lambda)f(\gamma(a)) + \lambda f(\gamma(b)), \forall \lambda \in [0, 1]$$

Definición 1.60. Sea M una variedad Riemanniana y $f : M \rightarrow \mathbb{R}$ una función real, f es llamada casi-convexa en M si para todo $x, y \in M, t \in [0, 1]$ se tiene que:

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

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

Teorema 1.7. Sea $f : M \rightarrow \mathbb{R}$ una función casi-convexa diferenciable en una variedad Riemanniana completa M y sean $x, y \in M$. Si $f(x) \leq f(y)$ entonces,

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

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

Demostración:

Como f es casi-convexa y además $f(x) \leq f(y)$ tenemos:

$$\begin{aligned} f(\gamma(t)) &\leq \max\{f(x), f(y)\}, \forall x, y \in M, t \in [0, 1] \\ &= f(y) \end{aligned}$$

Así, tenemos que:

$$\begin{aligned}f(\gamma(t)) \leq f(y) &\Rightarrow f(\gamma(t)) - f(y) \leq 0 \\&\Rightarrow \frac{f(\gamma(t)) - f(y)}{t} \leq 0, \forall t \in [0, 1] \\&\Rightarrow Df(y, \gamma'(0)) \leq 0 \\&\Rightarrow \langle \text{grad}f(y), \gamma'(0) \rangle \leq 0\end{aligned}$$

□

Capítulo 2

Desarrollo

2.1. Método de Cauchy

El Método de Cauchy , es uno de los métodos más antiguo y reconocidos en la literatura para resolver problemas de optimización. Sin embargo, los resultados clásicos de una función objetivo arbitraria, no son muy fuertes debido a que la convergencia plena y la existencia de puntos clausura no están asegurados. Solo se puede asegurar que cualquier punto clausura, si existe, es un punto critico del problema.

Por otro lado, la generalización de los espacios euclidianos a variedades Riemanniana tienen algunas ventajas importantes. Por ejemplo, problemas de optimización con restricciones se pueden ver sin restricciones desde el punto de vista de la geometría Riemanniana, otra ventaja es que los problemas de optimización con funciones objetivos no-convexas, se pueden convertir en convexas a través de la introducción de una adecuada métrica Riemanniana.

La motivación para estudiar este tema proviene de la amplia gama de aplicaciones de optimización casi-convexa en diversas áreas de las ciencias y la

ingeniería, en este contexto, si las restricciones de un problema de minimización constituyen una variedad Riemanniana, y la función objetivo es casi-convexo, el problema se convierte en un problema sin restricciones, y por lo tanto no es necesario hacer proyecciones en cada iteración del Método de Cauchy.

El interés de este trabajo es resolver problemas de optimización de la siguiente forma,

$$\min_{x \in M} f(x) \quad (2.1)$$

donde M es una variedad Riemanniana completa finito dimensional y $f : M \rightarrow \mathbb{R}$ una función casi-convexa continuamente diferenciable.

El método de Cauchy propuesto por [1] genera una sucesión $\{x^k\}$ que viene dada de la siguiente manera:

$$x^0 \in M \quad (2.2)$$

$$x^{k+1} = \exp_{x^k}(-t_k \text{grad} f(x^k)) \quad (2.3)$$

Donde \exp es la exponencial y t_k es algún tamaño de paso positivo.

En el caso de tener $M = \mathbb{R}^n$ (el espacio euclidiano) tenemos que (2.3) es equivalente a:

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

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

Existen diferentes maneras de escoger el parámetro t_k generando consecuentemente diversos submétodos, en este trabajo daremos dos de ellos.

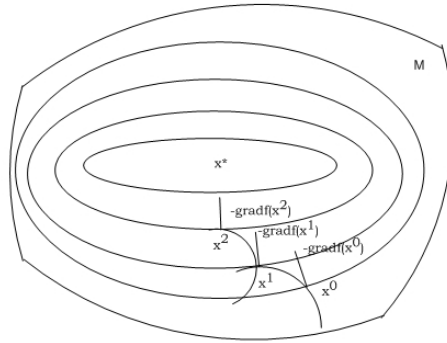


Figura 2.1: Esquema del proceso iterativo del método de Cauchy sobre una variedad M .

Asumiremos lo siguiente:

Hipótesis 1 (H1): El conjunto de óptimos globales del problema (2.1), denotado por X^* , es no vacío. Denotaremos el valor óptimo de (2.1) por f^* . Ahora, definiremos el siguiente conjunto no-vacío:

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

El siguiente lema, es la clave de este trabajo ya que se usará este hecho, para demostrar que la sucesión, definida por el método de Cauchy es casi-Fejér convergente en U .

Lema 2.1. *Sea $f : M \rightarrow \mathbb{R}$ una función casi-convexa continua y diferenciable en una variedad Riemanniana completa M 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 para todo $t_k > 0$.

Demostración. Sea $x \in U$ arbitrario, por definición del conjunto U . Entonces:

$$x \in M : f(x) \leq \inf_k f(x^k).$$

Supongamos que, $\gamma_1 : [0, 1] \rightarrow M$ es un segmento de geodésica minimal que conecta x^k con x , y $\gamma_2 : [0, 1] \rightarrow M$ un segmento de geodésica que conecta x^k con x^{k+1} con $\gamma_2'(0) = -t_k \text{grad}(x^k)$.

Por otra parte, tenemos que:

$$\ell_3^2 = \gamma_3^2 = d^2(x^{k+1}, x);$$

$$\ell_1^2 = \gamma_1^2 = d^2(x, x^k);$$

$$\ell_2^2 = \gamma_2^2 = d^2(x^k, x^{k+1});$$

$$-2\ell_1\ell_2 \cos \alpha = -2d(x, x^k)d(x^k, x^{k+1}) \cos \alpha;$$

$$2t_k d(x^k, x) \langle \text{grad}f(x^k), \gamma_1'(0) \rangle = 2t_k \|\text{grad}f(x^k)\| \|\gamma_1'(0)\| \cos \alpha = 2t_k d(x^k, x).$$

Aplicando el teorema (1.6) tenemos que:

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

Luego, como f es casi-convexa y $f(x) \leq f(x^k)$, por teorema (1.7) tenemos que, $\langle \text{grad}f(x^k), \gamma_1'(0) \rangle \leq 0$. Así,

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

□

De ahora en adelante M sera una variedad Riemanniana de dimensión finita con curvatura seccional no-negativa.

2.2. Método de Cauchy con búsqueda generalizada de Armijo

El Método de Cauchy con tamaño de paso de Armijo genera una sucesión $\{x^k\}$ propuesta por [1] dada por (2.2) y (2.3) donde:

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

En esta sección probaremos la convergencia plena de este método en el caso casi-convexo.

Consideremos las siguientes hipótesis:

Hipótesis (H2): Sea $\phi : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ una función tal que:

H2.1 Existe $\alpha \in (0, 1), \tau_\alpha > 0$, tal que, para todo $t \in (0, \tau_\alpha] : \phi(t) \leq \alpha t$.

H2.2 Existe $\beta > 0, \tau_\beta \in (0, +\infty]$, tal que, para todo $t \in (0, \tau_\beta] \cap \mathbb{R} : \phi(t) \geq \beta t^2$.

H2.3 Para todo $k, f(x^{k+1}) \leq f(x^k) - \phi(t_k) \|\operatorname{grad}f(x^k)\|^2$ con $0 < t_k \leq \tau_\beta$.

H2.4 Existe $\gamma > 1, \tau_\gamma > 0$, tal que: para todo $k : t_k \geq \tau_\gamma$ ó

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

Proposición 2.1. *Sea $f : M \rightarrow \mathbb{R}$ una función casi-convexa continuamente diferenciable. Supongamos que H1 y H2 se satisfacen. Entonces, la sucesión $\{x^k\}$ generada por el método de cauchy con la búsqueda generalizada de Armijo es casi-Fejér convergente en U .*

Demostración:

Como f es casi-convexa y continuamente diferenciable, por el teorema (2.1) tenemos que:

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

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

Queremos probar que la sucesión $\{x^k\}$ es casi-Fejér convergente en U , por definición debemos probar que:

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

Ahora, por las hipótesis H2.2 y H2.3 tenemos que:

$$\begin{aligned} f(x^{k+1}) &\leq f(x^k) - \beta t_k^2 \|\text{grad}f(x^k)\|^2 \Rightarrow \beta t_k^2 \|\text{grad}f(x^k)\|^2 \leq f(x^k) - f(x^{k+1}) \\ &\Rightarrow t_k^2 \|\text{grad}f(x^k)\|^2 \leq \frac{f(x^k) - f(x^{k+1})}{\beta}. \end{aligned}$$

Luego, esto implica que:

$$\begin{aligned} \sum_{k=0}^{+\infty} t_k^2 \|\text{grad}f(x^k)\|^2 &\leq \sum_{k=0}^{+\infty} \frac{f(x^k) - f(x^{k+1})}{\beta} \\ &\leq \frac{f(x^0) - f^*}{\beta} \\ &\leq +\infty. \end{aligned}$$

Por lo tanto, la sucesión $\{x^k\}$ es casi-Fejér convergente en U . \square

Teorema 2.1. *Sea $f : M \rightarrow \mathbb{R}$ una función casi-convexa continuamente diferenciable. Supongamos que H1 y H2 se satisfacen. Entonces, la sucesión $\{x^k\}$ generada por el método de cauchy con búsqueda generalizada de Armijo converge. Por otra parte, converge a un punto estacionario (un punto \bar{x} tal que $\text{grad}f(\bar{x}) = 0$).*

Demostración:

Por proposición anterior, la sucesión $\{x^k\}$ es casi-Fejér convergente en U , esto es, $\{x^k\}$ es cerrada (por teorema), entonces existe \bar{x} y una subsucesión $\{x^{k_j}\}$ de $\{x^k\}$ que converge a \bar{x} .

Por la continuidad de f , obtenemos:

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

Como $\{f(x^k)\}$ es una sucesión no-creciente ya que;

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

con una subsucesión que converge a $f(\bar{x})$ y por tanto,

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

Esto implica que, $\bar{x} \in U$ (por definición del conjunto U). Luego, por teorema podemos concluir que la sucesión $\{x^k\}$ converge a \bar{x} .

Finalmente probaremos que: $\text{grad}f(\bar{x}) = 0$. Supongamos por absurdo que $\text{grad}f(\bar{x}) \neq 0$.

Claramente, tenemos que:

$\text{grad}f(x^k) \rightarrow \text{grad}f(f(\bar{x})) \neq 0$ ya que $f(x^k) \rightarrow f(\bar{x})$ y el grad es una función continua.

Por un lado, tenemos que:

$$0 \leq \beta t_k^2 \|\text{grad}f(x^k)\|^2 \leq f(x^k) - f(x^{k+1}).$$

Luego, tomando límite cuando $k \rightarrow +\infty$ en la expresión anterior, tenemos:

$$\begin{aligned}
0 &\leq \beta t_k^2 \|\text{grad}f(\bar{x})\|^2 \leq f(\bar{x}) - f(x^k) = 0. \\
&\Rightarrow 0 \leq \beta t_k^2 \|\text{grad}f(\bar{x})\|^2 \leq 0.
\end{aligned}$$

Como $\beta > 0$, $\|\text{grad}f(\bar{x})\|^2 > 0$ entonces,

$$\lim_{k \rightarrow +\infty} t_k = 0. \quad (2.4)$$

Por otra parte, de H2.4 y H2.1 tenemos que para k suficientemente grande,

$$\begin{aligned}
f(\exp_{x^k}(-\bar{t}_k \text{grad}f(x^k))) &\geq f(x^k) - \phi(\bar{t}_k) \|\text{grad}f(x^k)\|^2 \\
&\geq f(x^k) - \alpha \bar{t}_k \|\text{grad}f(x^k)\|^2 \\
\Rightarrow f(\exp_{x^k}(-\bar{t}_k \text{grad}f(x^k))) - f(x^k) &\geq -\alpha \bar{t}_k \|\text{grad}f(x^k)\|^2 \quad (2.5)
\end{aligned}$$

Definamos, $g : [0, \bar{t}_k] \rightarrow \mathbb{R}$ tal que:

$$\begin{aligned}
g(t) &= f(\exp_{x^k}(-t \text{grad}f(x^k))), \\
g(0) &= f(\exp_{x^k}(0)) = f(x^k), \\
g(\bar{t}_k) &= f(x^{k+1}).
\end{aligned}$$

Así, por el teorema del Valor Medio, para tal k existe $t_k^* \in [0, \bar{t}_k]$ tal que:

$$f(x^{k+1}) - f(x^k) = g(\bar{t}_k) - g(0) = g'(t_k^*)(\bar{t}_k - 0) \geq \alpha \bar{t}_k \|\text{grad}f(x^k)\|^2$$

donde, $g'(t_k^*) = -\langle \text{grad}f(\exp_{x^k}(-t_k^* \text{grad}f(x^k))), \rho_{\gamma, 0, t_k^*} \text{grad}f(x^k) \rangle$, por (2.3).

Así,

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

donde $\rho_{\gamma_k, 0, t_k^*}$ 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, de (2.4) y de H2.4 implica que $\lim t_k^* = 0$.

Haciendo, $k \rightarrow +\infty$ en la desigualdad anterior y teniendo en cuenta la continuidad del $gradf, exp$ y del transporte paralelo, tenemos que:

$$-\frac{\|gradf(\bar{x})\|^2}{\|gradf(\bar{x})\|^2} \geq -\alpha \Rightarrow 1 \leq \alpha,$$

lo cual contradice H2.1

Por lo tanto, $gradf(\bar{x}) = 0$. □

2.3. El Método de Cauchy con una regularización Proximal, para el caso casi-convexo.

Sea $\{\lambda_k\}$ una sucesión tal que $\lambda' \leq \lambda_k \leq \lambda''$, donde $0 \leq \lambda' \leq \lambda''$. La regularización del método de Cauchy genera una sucesión $\{x^k\}$ definida por (2.2) y (2.3), tal que:

$$t_k = \operatorname{argmin}_{t>0} \varphi_k(t), \quad (2.6)$$

donde,

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

y sea;

$$x^{k+1} = \exp_{x^k}(-t_k gradf(x^k)) \quad (2.7)$$

Proposición 2.2. *Sea $\{x^k\}$ dada por (2.2), (2.3) y (2.6), esta bien definida y para todo k ,*

$$\langle gradf(x^{k+1}), P_{x^k, x^{k+1}} gradf(x^k) \rangle = 2\lambda_k t_k \|gradf(x^k)\|^2,$$

donde $P_{x^k, x^{k+1}}$ es el transporte paralelo a lo largo de $\exp_{x^k}(-t_k \text{grad}f(x^k))$ de x^k a x^{k+1} .

Demostración:

Supondremos que x^k es conocido. Tendremos dos casos.

- 1^{er} Caso: $\text{grad}f(x^k) = 0$

En este caso, $\varphi_k(t) = f(x^k)$ para todo t , y por lo tanto algún $t \geq 0$ es solución de (2.6) y por (2.7), $x^{k+1} = x^k$. En efecto:

$$x^{k+1} = \exp_{x^k}(-t_k \text{grad}f(x^k)) = \exp_{x^k}(0) = x^k$$

- 2^{do} caso: Supondremos que $\text{grad}f(x^k) \neq 0$, en este caso,

$$\varphi_k(t) \geq f^* + t^2 \lambda_k \|\text{grad}f(x^k)\|^2, \quad (2.8)$$

para todo $t \geq 0$. Tomando limite en (2.8) tenemos que:

$$\lim_{t \rightarrow \infty} \varphi_k(t) \geq \lim_{t \rightarrow \infty} (f^* + t^2 \lambda_k \|\text{grad}f(x^k)\|^2) = \infty.$$

De acá, $\lim_{t \rightarrow \infty} \varphi_k(t) = \infty$ entonces (2.6) tiene solución y x^{k+1} esta bien definida. Además, como $t_k = \text{argmin}_{t > 0} \varphi_k(t)$, entonces,

$$\begin{aligned} 0 &= \varphi'_k(t) \\ &= \langle \text{grad}f(\exp_{x^k}(-t \text{grad}f(x^k))), -\text{grad}f(x^k) \rangle + 2t \lambda_k \|\text{grad}f(x^k)\|^2, \end{aligned}$$

entonces,

$$\begin{aligned} \langle \text{grad}f(x^{k+1}), \text{grad}f(x^k) \rangle &= 2t_k \lambda_k \|\text{grad}f(x^k)\|^2 \\ \langle \text{grad}f(x^{k+1}), P_{x^k, x^{k+1}} \text{grad}f(x^k) \rangle &= 2t_k \lambda_k \|\text{grad}f(x^k)\|^2. \end{aligned}$$

□

Proposición 2.3. *Sea $f : M \rightarrow \mathbb{R}$ una función casi-convexa continuamente diferenciable. Supongamos que H1 se satisface. Entonces, la sucesión $\{x^k\}$ generada por (2.2), (2.3) y (2.6), es casi-Fejér convergente sobre el conjunto U .*

Demostración:

Como, $t_k = \operatorname{argmin}_{t>0} \varphi_k(t)$, donde:

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

Entonces, para cualquier $t \geq 0$ se cumple que: $\varphi_k(t_k) \leq \varphi_k(t)$. En particular, para $t = 0$, $\varphi_k(t_k) \leq f(x^k)$,

$$\begin{aligned} \Rightarrow f(x^{k+1}) &\leq f(x^k) + t_k^2 \lambda_k \|\operatorname{grad} f(x^k)\|^2 \\ &< f(x^k). \end{aligned}$$

Así, $\{f(x^k)\}$ es una sucesión no-creciente. Y además,

$$\begin{aligned} t_k^2 \lambda_k \|\operatorname{grad} f(x^k)\|^2 &\leq f(x^k) - f(x^{k+1}) \\ \Rightarrow \|\operatorname{grad} f(x^k)\|^2 &\leq \frac{f(x^k) - f(x^{k+1})}{t_k^2 \lambda_k} \end{aligned}$$

De aquí tenemos que,

$$\begin{aligned} \sum_{k=0}^{+\infty} t_k^2 \|\operatorname{grad} f(x^k)\|^2 &\leq \sum_{k=0}^{+\infty} t_k^2 \left(\frac{f(x^k) - f(x^{k+1})}{t_k^2 \lambda_k} \right) \\ &\leq \sum_{k=0}^{+\infty} \frac{f(x^k) - f(x^{k+1})}{\lambda_k} \\ &\leq \sum_{k=0}^{+\infty} \frac{1}{\lambda'} (f(x^k) - f(x^{k+1})) \\ &= \frac{1}{\lambda'} (f(x^0) - f(x^*)) \\ &< +\infty. \end{aligned}$$

Por lo tanto, la sucesión $\{x^k\}$, generada por (2.2), (2.3) y (2.6), es casi-Fejér convergente. Además, la serie $\sum_{k=0}^{+\infty} t_k^2 \|\text{grad}f(x^k)\|^2$ converge y por teorema sabemos que $t_k^2 \|\text{grad}f(x^k)\|^2 \rightarrow 0$. \square

Teorema 2.2. *Sea $f : M \rightarrow \mathbb{R}$ una función casi-convexa continuamente diferenciable. Supongamos que H1 se satisface. Entonces, la sucesión $\{x^k\}$ generada por (2.2), (2.3) y (2.6), converge a un punto estacionario.*

Demostración:

De las proposiciones (2.2) y (2.3) tenemos que:

- $\{f(x^k)\}$ es una sucesión no-creciente.
- La serie $\sum_{k=0}^{+\infty} t_k^2 \|\text{grad}f(x^k)\|^2$ converge y por teorema sabemos que: $t_k^2 \|\text{grad}f(x^k)\|^2 \rightarrow 0$.
- $\langle \text{grad}f(x^{k+1}), P_{x^k, x^{k+1}} \text{grad}f(x^k) \rangle = 2\lambda_k t_k \|\text{grad}f(x^k)\|^2$.

Sea \bar{x} un punto clausura de la sucesión $\{x^k\}$ y $\{x^{kj}\}$ una subsucesión de $\{x^k\}$ que converge a \bar{x} , además sabemos que la sucesión $\{x^{kj+1}\}$ converge también a \bar{x} .

Por ser la métrica Riemanniana, el transporte paralelo y el gradiente funciones continuas, entonces:

$$\lim_{j \rightarrow +\infty} \langle \text{grad}f(x^{kj+1}), P_{x^{kj}, x^{kj+1}} \text{grad}f(x^{kj}) \rangle = \|\text{grad}f(\bar{x})\|^2$$

Luego, por el ítem 2 y 3, se tiene:

$$\|\text{grad}f(\bar{x})\|^2 = \lim_{j \rightarrow +\infty} 2\lambda_{kj} t_{kj} \|\text{grad}f(x^{kj})\|^2 = 0$$

$$\Rightarrow \|gradf(\bar{x})\|^2 = 0$$

$$\Rightarrow \|gradf(\bar{x})\| = 0$$

Por lo tanto, la sucesión $\{x^k\}$ generada por (2.2), (2.3) y (2.6) converge a un punto estacionario \bar{x} . \square

Conclusión

En este trabajo se describió el resultado de convergencia plena del Método de Cauchy para resolver problemas de minimización de funciones objetivos casi-convexas en Variedades Riemannianas propuesto por [1] con una búsqueda generalizada de Armijo y una regularización proximal. Donde podemos notar que estos métodos resuelve los problemas de minimización con funciones objetivos convexas en el espacio Euclideo, si las restricciones constituyen una Variedad Riemanniana finito dimensional con curvatura seccional no negativa y la función objetivo es casi-convexa en esa Variedad.

Bibliografía

- [1] E.A. Papa Quiroz E.M. Quispe, P. Roberto Oliveira. Steepest descent method with a generalized armijo search for quasiconvex functions on riemannian manifolds. *Departament of Systems Engineering and Computer Science*, pages 467–477, 2007.
- [2] D. M.P. doCarmo. *Riemannian Geometry*. Birkhauser, Boston, Nashua, USA, 1992.