
OPTIMIZACIÓN EN VARIEDADES

PEDECIBA

Universidad de la República

15 de marzo de 2018

Comentarios

Estas notas fueron realizadas para el curso *Optimización en Variedades* dictado por Diego Armentano en el segundo semestre de 2017. Este curso se enmarcó dentro de los cursos de posgrado de PEDECIBA Matemática, y también válido para la carrera de Ingeniería Matemática.

Las referencias principales para estas notas fueron los libros de Helmke y Moore [HM], y el de Absil, Mahony y Sepulchre [AMS]. También se utilizó en varias partes el artículo Edelman, Arias y Smith [EAS].

Se utilizó material extra que se puede encontrar en las referencias, mucho de los cuales sirvieron para hacer el apéndice.

Las notas fueron tipeadas casi en su totalidad por el gran Federico Carrasco. Así que cualquier error existente es responsabilidad de él ;)

También colaboraron en la redacción de estas notas: Matías Valdés (Sección 4.1), Marcelo Fiori (Sección 4.3), Leandro Bentancur (Capítulo 6), Jasmín Finot (Capítulo 7), y Florencia Cubría (Capítulo D del apéndice),

Índice general

1. Introducción	9
1.1. Descripción del curso	10
I Métodos globales	13
2. Algoritmos clásicos para valores propios	15
2.1. Método de la Potencia	15
2.1.1. Método de la potencia sobre $\mathbb{C}P^{n-1}$	17
2.1.2. Método de la potencia vs ecuación de Riccati	18
2.1.3. Subespacios invariantes	21
2.1.4. Método de la potencia en $\text{Flag}(\mathbb{C}^n)$	23
2.2. Cociente de Rayleigh	26
2.2.1. Flujo gradiente de Rayleigh	27
2.3. Cociente de Rayleigh generalizado	29
2.3.1. Geometría de la variedad de Stiefel	30
2.3.2. Aspectos geométricos y dinámicos del flujo	34
2.4. Algoritmo QR	35

2.4.1. Algoritmo QR y flujo asociado	38
2.5. Descomposición en valores singulares	40
2.5.1. Flujos isosingulares	41
2.6. Mínimos cuadrados	41
3. Flujos Isoespectrales	43
3.1. Flujo en $\mathcal{O}(n)$	48
3.2. Flujos Toda	50
3.3. Flujos isoespectrales y ecuación de Riccati	54
4. Otras aplicaciones	57
4.1. Programación lineal en el caso convexo	57
4.1.1. Ecuación de Brockett	58
4.1.2. Ecuación de Brockett simplificada (en la esfera)	59
4.1.3. Ecuación de Brockett simplificada (para PL en el simplex)	61
4.2. Programación Lineal en un Polítopo	62
4.2.1. Gradiente de ϕ en \mathring{C}	63
4.2.2. Flujo de gradiente de ϕ en \mathring{C}	64
4.2.3. Flujo en el Simplex	64
4.3. Flujos en $\mathcal{O}(n)$ y Graph Matching	65
II Métodos locales	71
5. Retracciones y búsqueda lineal	73
5.1. Ejemplos de retracciones	74
5.1.1. Retracciones en variedades encajadas	75
5.2. Retracciones en variedades cociente	76

5.3. Algoritmos de búsqueda lineal	78
6. Geometría diferencial computacional	81
6.1. Variedad de Stiefel euclídea	82
6.1.1. Geodésicas en $St(n, p)$	82
6.2. Variedad de Stiefel cociente	85
6.3. Geometría de Grass(p, n)	87
6.3.1. Grass(p, n) como cociente de $St(p, n)$	89
6.4. Hessiano de funciones	90
6.4.1. Hessiano de una función en $St(p, n)$ y Grass(p, n)	90
7. Método de Newton	93
7.1. Método de Newton en \mathbb{R}^n	93
7.2. Método de Newton en variedades	95
7.2.1. Método de Newton para puntos críticos	97
7.2.2. Cociente de Rayleigh en la esfera	97
7.2.3. Método de Newton en Grass(p, n)	98
III Apéndice	101
A. Topología	103
A.1. Relaciones de equivalencia.	103
A.2. Espacios topológicos	104
A.3. Funciones continuas y homeomorfismos	105
A.4. Bases	106
A.5. Espacio producto	106
A.6. Espacios métricos	107

A.6.1. Topología métrica	109
A.7. Completitud	109
A.8. Compacidad	109
A.9. Topología cociente	111
B. Geometría diferencial	113
B.1. Variedad diferenciable	113
B.1.1. Mapas diferenciables	114
B.1.2. Vectores tangentes y espacio tangente	115
B.1.3. Cambio de coordenadas	117
B.2. Fibrado tangente	118
B.3. Campo de vectores	118
B.3.1. Corchete de Lie	120
B.3.2. Ecuaciones diferenciales y subgrupos a un parámetro	120
B.4. Subvariedades	122
B.5. Puntos críticos	123
B.6. Hessiano	124
B.7. Funciones de Morse, Morse-Bott	126
B.8. Métrica Riemanniana	127
B.8.1. Métrica inducida por inmersión	127
B.9. Gradiente	128
B.10. Flujo gradiente	129
C. Grupos de Lie	133
C.1. Espacio tangente a un grupo de Lie	135
C.2. Subgrupos a un parámetro	136
C.3. Acciones de grupos	138

C.4. Espacios homogéneos	140
C.5. Métricas Riemannianas en grupos de Lie	144
C.5.1. Métricas bi-invariantes	144
C.6. Métricas Riemannianas en espacios homogéneos	145
D. Conexiones	147
D.1. Introducción	147
D.2. Conexiones afines	147
D.3. Conexión Riemanniana	151
D.4. Conexiones en subvariedades Riemannianas	153
D.4.1. Ejemplos	154
D.5. Conexión Riemanniana en espacios homogéneos	155
D.6. Campo aceleración	156
D.7. Hessiano	159

Introducción

Optimización en variedades es un área que se encuentra en pleno desarrollo motivado por sus múltiples aplicaciones en distintas disciplinas científicas. Básicamente el objetivo es el siguiente.

Dada M una variedad, y $f : M \rightarrow \mathbb{R}$ una función diferenciable, encontrar $x_ \in M$ tal que $f(x_*) \leq f(x)$ para x en un entorno de x_* en M . (i.e. x_* es un mínimo local de f .)*

Un algoritmo que intente resolver este problema producirá una sucesión x_0, x_1, \dots en M que se aproxime a x_* .

Hay al menos dos escenarios distintos que motivan este problema. El primero consiste en problemas de optimización clásicos de funciones definidas en \mathbb{R}^n sujetas a *restricciones* (por ejemplo tomando M la esfera S^{n-1}). El segundo escenario surge de considerar funciones definidas en \mathbb{R}^n que tienen *simetrías*, o son *invariantes* bajo cierto grupo de transformaciones. Un ejemplo importante es cuando consideramos funciones que son constantes en las órbitas de una acción de un grupo G sobre \mathbb{R}^n . En tal caso, un espacio natural para trabajar es el espacio de “órbitas”, (i.e. el espacio cociente $M = \mathbb{R}^n/G$). Ejemplos clásicos son espacios proyectivos, grassmanianas, espacios homogéneos en general.)

Históricamente este tipo de problemas, con restricciones en el dominio, se atacaban con técnicas de *optimización no-lineal* en \mathbb{R}^n . Sin embargo, en las últimas décadas se observó que el análisis de este problema de optimización de manera “intrínseca” a la variedad resultó en el diseño de algoritmos más eficientes. (Este nuevo enfoque refuerza el principio general en matemática de que para estudiar un problema es recomendable primero eliminar las redundancias.)

En este curso pondremos especial énfasis a problemas clásicos de álgebra lineal como el problema de valores propios, valores singulares, o mínimos cuadrados.

Es interesante ver como estos problemas básicos motivan la utilización de herramientas provenientes de distintas áreas de la matemática, como álgebra, análisis, geometría, y sistemas dinámicos.

1.1. Descripción del curso

En la primera parte del curso veremos cómo problemas clásicos de álgebra lineal pueden ser abordados con implementaciones de sistemas dinámicos (continuos y discretos). En particular, especial énfasis se dará al estudio de flujos gradiente, y flujos “isoepectrales”, y discretizaciones de los mismos.

En esta primer parte los problemas que estudiaremos son los relacionados a encontrar *vectores propios dominantes* y *subespacios dominantes*.

En otro capítulo estudiaremos el problema de *programación lineal*, y también problemas clásicos relacionados a *clasificación de grafos*.

En la segunda parte estudiaremos métodos que hagan uso de la geometría “de segundo orden” (i.e. conexiones afines y riemannianas). En particular se estudiará *retracciones*, y el *método de Newton* en variedades Riemannianas. Para tal fin nos aproximaremos a la geometría diferencial desde un punto de vista computacional. Este enfoque será de utilidad para implementaciones de algoritmos.

Hemos agregado un apéndice que contiene la gran mayoría de los requisitos técnicos de geometría diferencial. Algunos de estos temas no fueron abordados durante el curso, pero sí se dedico buena parte del curso al estudio de *grupos de Lie* y *espacios homogéneos* dado que estos son los espacios que más nos interesa

para las aplicaciones.

También se dedicó un tiempo a estudiar conexiones en variedades, que sirven como una herramienta básica para poder estudiar algoritmos más finos y rápidos que los estudiados en la primer parte del curso.

Parte I

Métodos globales

Algoritmos clásicos para valores propios

2.1. Método de la Potencia

Sea $A : \mathbb{C}^m \rightarrow \mathbb{C}^m$ una matriz diagonalizable con espectro $Sp(A) = \{\lambda_1, \dots, \lambda_m\}$, donde los ordenamos por

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_m|,$$

y sean v_1, \dots, v_m los vectores propios correspondientes. En este caso, decimos que v_1 es *vector propio dominante*.

PROBLEMA (Vector propio dominante): *¿Cómo encontrar el vector propio dominante de A ?*

El *método de la potencia* es el algoritmo más sencillo para resolver este problema. Su belleza y simplicidad, desde el punto de vista de sistemas dinámicos, lo hace muy atractivo. Sin embargo este método tiene un lado negativo, no es eficiente desde el punto de vista del análisis numérico.

En esta sección analizaremos este método y lo relacionaremos con la ecuación de Riccati.

Sobre \mathbb{C}^m consideramos la estructura hermitiana usual. Al producto hermitiano lo denotamos por $\langle \cdot, \cdot \rangle$, y por $\| \cdot \|$ a la norma.

El *método de la potencia* se define por la iteración siguiente

$$x_{k+1} := \frac{Ax_k}{\|Ax_k\|}, \quad k \in \mathbb{N},$$

donde $x_0 \in \mathbb{C}$ es cierto vector unitario inicial. (Observar que el método está definido siempre que $Ax_k \neq 0$).

Tomando como base del espacio los vectores propios de A resulta natural que la sucesión definida anteriormente converja a ηv_1 , para cierto $\eta \in \mathbb{C}$, con $|\eta| = 1$. Sin embargo, como veremos a continuación, esto no siempre es verdad.

Ejemplo 2.1. Sea $A = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, y $x_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Resulta

$$x_k = \frac{1}{\sqrt{1+2^{2k}}} \begin{pmatrix} 1 \\ 2^k \end{pmatrix},$$

de donde resulta

$$\lim_k x_k = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Ejemplo 2.2. Sea $A = \begin{pmatrix} 1 & 0 \\ 0 & -2 \end{pmatrix}$, y $x_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Resulta

$$x_k = \frac{1}{\sqrt{1+2^{2k}}} \begin{pmatrix} 1 \\ (-1)^k 2^k \end{pmatrix},$$

de donde resulta x_k acumula en el conjunto

$$\left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ -1 \end{pmatrix} \right\},$$

y por tanto su límite no está definido.

◇ **2.3** (Dinámica en S^1). Extender los ejemplos anteriores, en el caso real, a condiciones iniciales $x_0 \in S^1 = \{x \in \mathbb{R}^2 : \|x\| = 1\}$. Estudiar el comportamiento dinámico de la sucesión x_k en S^1 .

De lo anterior surge, que aunque la sucesión x_k no tenga límite, sus puntos de acumulación pertenecen al subespacio propio asociado al vector dominante. Esto motiva a definir el método de la potencia sobre el espacio proyectivo $\mathbb{R}\mathbb{P}^{n-1}$ o $\mathbb{C}\mathbb{P}^{n-1}$.

Disgresión: sea V un \mathbb{R} -espacio vectorial. Denotaremos por $\mathbb{P}(V)$ al *espacio proyectivo asociado*. Para el caso de \mathbb{C}^n usaremos indistintamente la notación $\mathbb{P}(\mathbb{C}^{n+1}) = \mathbb{C}\mathbb{P}^n$. Idem para el caso real.

2.1.1. Método de la potencia sobre $\mathbb{C}\mathbb{P}^{n-1}$

Una transformación lineal $A : \mathbb{C}^n \rightarrow \mathbb{C}^n$, invertible, induce un mapa

$$\widehat{A} : \mathbb{C}\mathbb{P}^{n-1} \rightarrow \mathbb{C}\mathbb{P}^{n-1},$$

de manera natural: $\widehat{A}(\ell) = A\ell$, donde $A\ell$ denota el subespacio unidimensional generado por la imagen de ℓ por A .

Definimos en este caso el *método de la potencia* como la sucesión

$$\ell_{k+1} := \widehat{A}(\ell_k), \quad k \in \mathbb{N}.$$

Teorema 2.4. Si A diagonalizable, con $Sp(A) = \{\lambda_1, \dots, \lambda_m\}$, ordenados por

$$|\lambda_1| > |\lambda_2| \geq \dots \geq |\lambda_m|,$$

Entonces, si ℓ_0 es genérico, y $\ell_k = \widehat{A}^k(\ell_0)$ se tiene

$$\lim_k \ell_k = [v_1],$$

siendo $[v_1]$ el subespacio generado por el vector propio dominante.

Demostración. Dado que A es diagonalizable, existe $X \in \mathbb{G}l(n, \mathbb{C})$, y D diagonal tal que $A = XDX^{-1}$. Por lo tanto los mapas \widehat{A} y \widehat{D} son *topológicamente conjugados* por el homeomorfismo $\widehat{X} : \mathbb{C}\mathbb{P}^{n-1} \rightarrow \mathbb{C}\mathbb{P}^{n-1}$. (En particular, $(\widehat{A})^n = \widehat{X}(\widehat{D})^n(\widehat{X})^{-1}$, $n \in \mathbb{N}$, donde $(\widehat{X})^{-1} = \widehat{X}^{-1}$.) Resulta entonces que basta probar el teorema para el caso $A = \text{diag}(\lambda_1, \dots, \lambda_n)$.

Sea $\mathcal{U}_1 \subset \mathbb{C}\mathbb{P}^{n-1}$ el abierto dado por $\ell_0 = [x_1 : \dots : x_n]$, tales que $x_0 \neq 0$.

Se tiene

$$\begin{aligned} (\widehat{A})^k \ell_0 &= (\widehat{A})^k [1 : y_2 : \dots : y_n] \\ &= [\lambda_1^k : \lambda_2^k y_2 : \dots : \lambda_n^k y_n] \\ &= [1 : (\lambda_2/\lambda_1)^k y_2 : \dots : (\lambda_n/\lambda_1)^k y_n] \end{aligned}$$

Dado que $(\lambda_i/\lambda_1)^k \rightarrow 0$, para todo $i = 2, \dots, n$, resulta que el $(\widehat{A})^k \ell_0 \rightarrow [1 : 0 : \dots : 0]$.
¹ □

◇ **2.5.** Generalizar el resultado anterior para el caso en que ℓ_0 pertenezca al subespacio propio generado por vectores propios.

2.1.2. Método de la potencia vs ecuación de Riccati

La idea de esta sección es estudiar el método de la potencia a través de una carta local. Para simplificar, tomemos el abierto \mathcal{U}_1 y la carta local ϕ_1 .

Consideremos la descomposición en bloques $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, donde $A_{11} \in \mathbb{C}^{1 \times 1}$, $A_{22} \in \mathbb{C}^{(n-1) \times (n-1)}$.

Dado $\ell_i = [1 : K_i] \in \mathcal{U}_1$, y asumiendo $\widehat{A}\ell_i \in \mathcal{U}_1$, de la igualdad

$$\begin{bmatrix} 1 \\ K_{i+1} \end{bmatrix} = \widehat{A} \begin{bmatrix} 1 \\ K_i \end{bmatrix}, \quad (2.1)$$

resulta

$$K_{i+1} = \frac{A_{21} + A_{22}K_i}{A_{11} + A_{12}K_i}, \quad (2.2)$$

definiendo de esta manera una sucesión $\{K_i\}_{i \in \mathbb{N}}$ en \mathbb{C}^{n-1} .

De esta manera tenemos que

$$\phi_1(\ell_i) = K_i, \quad i \in \mathbb{N}, \quad (2.3)$$

siempre que $\ell_i \in \mathcal{U}_1$.

La elección de \mathcal{U}_1 es completamente arbitraria.

¹Una forma de concluir este límite es hacerlo en cartas locales, i.e., si consideramos la carta local asociada al abierto \mathcal{U}_1 , a saber, $\phi_1 : \mathcal{U}_1 \rightarrow \mathbb{C}^n$, dada por $\phi_1([x_1 : \dots : x_n]) = (x_2/x_1, \dots, x_n/x_1)$, se tiene que $\phi_1(\widehat{A}^k(\ell_0)) \rightarrow 0$.

◇ **2.6.** Describir el conjunto de $\ell \in \mathbb{C}P^{n-1}$ tales que $\widehat{A}\ell \notin \mathcal{U}_1$. Probar que el conjunto de $\ell \in \mathbb{C}P^n$ tales que la iteración (2.1) no está definida, tiene medida de cero en $\mathbb{C}P^n$.

De (2.2) resulta

$$K_{i+1} - K_i = \frac{1}{A_{11} + A_{12}K_i} [A_{21} + A_{22}K_i - A_{11}K_i - K_i A_{12}K_i].$$

Esta identificación sugiere que la sucesión K_i es una aproximación discreta de la ecuación diferencial de Riccati (vectorial) en \mathbb{C}^{n-1}

$$\dot{K}(t) = \frac{d}{dt}K(t) = A_{21} + A_{22}K(t) - A_{11}K(t) - K(t)A_{12}K(t). \quad (2.4)$$

con condición inicial $K_0 = \phi_1(\ell_0)$.

Observar que puntos fijos de la sucesión K_i definida en (2.3) se corresponden con puntos de equilibrio de la ecuaciones diferencial (2.4). Además, es fácil ver que los subespacios propios se corresponden. Por lo tanto, estos puntos se corresponden con las soluciones algebraicas del sistema (a lo sumo cuadrático)

$$A_{21} + A_{22}K - A_{11}K - KA_{12}K = 0, \quad K \in \mathbb{C}^{n-1} \quad (2.5)$$

Comentario 2.7. La sucesión (2.3) está definida aunque la matriz no sea diagonalizable (a menos de los ℓ problemáticos de ◇2.6).

Comentario 2.8. Podemos entender las soluciones de la ecuación de Riccati mediante la identificación con el método de la potencia. Si estamos parados en las buenas coordenadas, y la matriz resulta diagonalizable con valores propios de diferente módulo, entonces el sistema (2.5) tiene n soluciones. Observar que el sistema de ecuaciones anterior tienen $n - 1$ variables y de grado 2 (genéricamente).

◇ **2.9.** Mostrar con un ejemplo que la sucesión $\{K_i\}$ puede no estar acotada.

En los siguiente párrafos daremos una prueba formal de las discusiones anteriores.

Dada una matriz $X \in \mathbb{C}^{n \times n}$ definimos el *mapa exponencial*

$$\exp: \mathbb{C}^{n \times n} \rightarrow \text{Gl}(n, \mathbb{C}), \quad \exp(X) = \sum_{n=0}^{+\infty} \frac{X^n}{n!}.$$

(Usaremos indistintamente la notación $e^X = \exp(X)$.)

Recordar que la solución de la ecuación diferencial matricial $\dot{\alpha} = X\alpha$, con condición inicial $\alpha(0) = C \in \mathbb{C}^{n \times n}$, está dada por $\alpha(t) = Ce^{tX}$, $t \in \mathbb{R}$.

Proposición 2.10. *Sea $F \in \mathbb{C}^{n \times n}$, y consideremos la partición en bloques $F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix}$ donde $F_{11} \in \mathbb{C}^{1 \times 1}$ y $F_{22} \in \mathbb{C}^{(n-1) \times (n-1)}$.*

Dado $\ell_ \in \mathbb{C}\mathbb{P}^{n-1}$, sea $\ell(t) = e^{tF}\ell_*$, con $t \in \mathbb{R}$. Si escribimos dicha curva en coordenadas homogéneas $\ell(t) = [\ell_0(t) : \dots : \ell_{n-1}(t)]$, y $\ell_0(t) \neq 0$, entonces la curva $K(t) = (\ell_1(t)/\ell_0(t), \dots, \ell_{n-1}(t)/\ell_0(t)) \in \mathbb{C}^{n-1}$ es solución de la ecuación de Riccati*

$$\dot{K} = F_{21} + F_{22}K - F_{11}K - KF_{12}K \tag{2.6}$$

con condición inicial $\phi_1(\ell_*)$.

◇ **2.11.** Probar el resultado anterior.

Corolario 2.12. *Hay una correspondencia biyectiva entre soluciones de la ecuación de Riccati (2.6) y las soluciones de la ecuación diferencial lineal $\dot{X} = FX$, con $X \in \mathbb{C}^{n \times n}$.*

Esta correspondencia es la que ayuda a resolver la ecuación de Riccati en una variable.

Proposición 2.13. *Sea $x(t)$ la solución de la ecuación diferencial $\dot{x} = ax^2 + bx + c$ con $a \neq 0$ (ecuación diferencial de Riccati). Entonces, si $\dot{y} = -axy$, (i.e. $y(t) = \exp(\int_0^t x(y) du)$), resulta $\ddot{y}(t) = b\dot{y} - acy$. Recíprocamente, si $y(t)$ satisface $\dot{x}(t) = \dot{y}(t)/y(t)$, con $y(t) \neq 0$ satisface la ecuación de Riccati*

◇ **2.14.** Dar una prueba de la proposición anterior y vincular con el **Corolario 2.12**.

Definición 2.15 (Logaritmo). Una matriz $A \in \mathbb{G}l(n, \mathbb{C})$ tiene un *logaritmo* si existe una matriz $\log A$ tal que $e^{\log A} = A$.

Si A tiene un logaritmo, a saber, $F = \log A$, se tiene que $A^k = e^{kF}$, y por lo tanto las sucesiones $\{(\widehat{A})^k \ell_0\}$ y $\{(e^{\widehat{F}})^k \ell_0\}$ coinciden. Por lo tanto podemos concluir lo siguiente.

Proposición 2.16. *El método de la potencia para una matriz A que tenga logaritmo se corresponde con muestras en tiempo natural de la solución de Riccati (2.6) para $F = \log A$.*

Una pregunta razonable es saber cuándo existe un logaritmo de matriz.

◇ **2.17.** Probar que para toda matriz $A \in \mathbb{G}l(n, \mathbb{C})$ diagonalizable se tiene que existe un logaritmo.

Cuando A tiene más estructura, por ejemplo hermitiana o simétrica, no siempre existe un logaritmo con las mismas propiedades. Por ejemplo en el caso simétrico cuando algún valor propio es negativo es fácil ver que no puede tener un logaritmo. Sin embargo para nuestros propósitos podemos decir lo siguiente.

◇ **2.18.** Si $A \in \mathbb{R}^{n \times n}$ es simétrica, entonces siempre existen matrices simétricas S y F tales que $A = Se^F$, donde $S^2 = \text{Id}_n$.

En este último caso, la correspondencia de la Proposición 2.16 es en los tiempos pares.

2.1.3. Subespacios invariantes

Una forma natural de extender el problema anterior es preguntarse lo siguiente.

PROBLEMA (Subespacios invariantes): Dada $A \in \mathbb{C}^{n \times n}$, ¿cómo encontramos el subespacio de dimensión k dominante?

El espacio natural para mirar este problema es la $\text{Grass}_{\mathbb{C}}(n, k)$, y considerar sobre este espacio la dinámica inducida por A . Esto es, dada $A \in \mathbb{G}l(n, \mathbb{C})$, sea

$$\hat{A} : \text{Grass}_{\mathbb{C}}(n, k) \rightarrow \text{Grass}_{\mathbb{C}}(n, k), \quad V \mapsto A \cdot V,$$

donde se entiende por $A \cdot V$ al subespacio imagen de aplicar A al subespacio $V \subset \mathbb{C}^n$.

Teorema 2.19. *Si asumimos $|\lambda_1| \geq \dots \geq |\lambda_k| > |\lambda_{k+1}| \geq \dots \geq |\lambda_n|$. Para todo $V \in \text{Grass}_{\mathbb{C}}(n, k)$ en un abierto y denso, la sucesión $(\hat{A})^r(V)$ es convergente al subespacio dominante.*

Demostración. Al igual que hicimos en el caso del proyectivo, podemos suponer que la matriz A es diagonal, de la forma $A = \text{diag}(\Lambda_1, \Lambda_2)$, donde $\Lambda_1 = \text{diag}(\lambda_1, \dots, \lambda_k)$, y $\Lambda_2 = \text{diag}(\lambda_{k+1}, \dots, \lambda_n)$.

Dada una matriz $X \in \mathbb{C}^{n \times k}$ de rango máximo, sea $[X] \in \text{Grass}_{\mathbb{R}}(n, k)$ el subespacio k -dimensional generado por las columnas de X . Sea $V = \left[\begin{array}{c} X_1 \\ X_2 \end{array} \right] \in \text{Grass}_{\mathbb{C}}(n, k)$, para $X_1 \in \mathbb{C}^{k \times k}$ invertible, y $X_2 \in \mathbb{C}^{(n-k) \times k}$.

Luego se tiene

$$(\hat{A})^r V = \left[\begin{array}{c} \Lambda_1^r X_1 \\ \Lambda_2^r X_2 \end{array} \right] = \left[\begin{array}{c} \Lambda_1^r X_1 \\ \Lambda_2^r X_2 \end{array} \right] (\Lambda_1 X_1)^{-r} = \left[\begin{array}{c} \text{Id}_k \\ \Lambda_2^r X_2 X_1^{-1} \Lambda_1^{-r} \end{array} \right].$$

Observar que $\|\Lambda_2^r X_2 X_1^{-1} \Lambda_1^{-r}\| \leq \|\Lambda_2\|^r \|X_2 X_1\| \|\Lambda_1\|^{-r} \leq (|\lambda_{k+1}|^r / |\lambda_k|^r) \|X_2 X_1\|$, y por lo tanto tiende a cero, cuando r tiende infinito.

Sólo resta justificar que sobre $\text{Grass}_{\mathbb{C}}(n, k)$ esto significa que $\hat{A}^r V$ tiende a $\left[\begin{array}{c} \text{Id}_k \\ 0_{(n-k) \times k} \end{array} \right]$. Esto resulta de que $\text{Grass}_{\mathbb{C}}(n, k)$ puede obtenerse como un cociente de las matrices de rango k , cocientando por la acción de $\text{Gl}(k, \mathbb{C})$ a derecha: esto es, $X \sim Y$ si existe $U \in \text{Gl}(k, \mathbb{C})$ tal que $Y = XU$. Luego, dado que la proyección cociente es continua se tiene que si una sucesión es convergente en el espacio de matrices de rango máximo, entonces su proyección en el cociente también es convergente. Esto termina el argumento de convergencia.

Para terminar la demostración hay que probar que el conjunto de V 's dada forma un abierto y denso. Esto queda de ejercicio. \square

◇ **2.20.** Terminar la prueba del teorema anterior.

Al igual que antes, queremos relacionar el método de la potencia para encontrar el k -subespacio dominante con la ecuación de Riccati para $F = \log A$.

Sea

$$F = \begin{pmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{pmatrix},$$

donde $F_{11} \in \mathbb{C}^{k \times k}$, $F_{22} \in \mathbb{C}^{(n-k) \times (n-k)}$.

Si $K(t)$ es una solución de la ecuación de Riccati (2.6) en $\mathbb{C}^{(n-k) \times k}$, entonces

la curva

$$V(t) = \left[\begin{array}{c} \text{Id}_k \\ K(t) \end{array} \right] \in \text{Grass}_{\mathbb{C}}(n, k),$$

y resulta

$$V(t) = e^{tF} V(0), \quad \forall t \in \mathbb{R}.$$

Recíprocamente, si $V(t) = e^{tF} V(0)$, y escribimos

$$V(t) = \left[\begin{array}{c} X_1(t) \\ X_2(t) \end{array} \right],$$

entonces, $K(t) = X_2(t)X_1(t)^{-1}$ es solución de la ecuación de Riccati (2.6), siempre que $X_1(t)$ sea invertible.

◇ **2.21.** Probar que el método de la potencia para encontrar el k -subespacio dominante coincide con muestras enteras de la solución de la ecuación de Riccati (2.6) para $F = \log(A)$.

2.1.4. Método de la potencia en $\text{Flag}(\mathbb{C}^n)$

Una bandera es una sucesión creciente de subespacios $V_0 \subset V_1 \subset \dots \subset V_k \subset \mathbb{R}^n$. Nosotros estudiaremos el caso de las banderas completas, donde $n = k$ y $\dim V_i = i$.

Definición 2.22 (Espacio bandera). Una *bandera* es una $n + 1$ -úpla

$$F = (F^{(0)}, \dots, F^{(n)})$$

donde los $F^{(i)}$ son subespacios vectoriales de \mathbb{R}^n tales que $F^{(i)} \subset F^{(i+1)}$ y $\dim F^{(i)} = i$. En particular $F^{(0)} = \{0\}$ y $F^{(n)} = \mathbb{R}^n$. Denotaremos por $\text{Flag}(\mathbb{R}^n)$ al espacio de banderas de \mathbb{R}^n . Análogo si cambiamos \mathbb{R}^n por \mathbb{C}^n , en tal caso lo denotaremos $\text{Flag}(\mathbb{C}^n)$.

Observación 2.23. Dado $F \in \text{Flag}(\mathbb{C}^n)$ existe $X = (X_1, \dots, X_n) \in \mathbb{G}l(n, \mathbb{C})$ tal que $F^{(i)} = \text{span}\{X_1, \dots, X_i\}$. Además, si $Y = (Y_1, \dots, Y_n) \in \mathbb{G}l(n, \mathbb{C})$ es otra matriz con la misma propiedad, entonces $Y = XR$ con R triangular superior invertible. Luego podemos identificar $\text{Flag}(\mathbb{C}^n)$ como el cociente

$$\text{Flag}(\mathbb{C}^n) = \mathbb{G}l(n, \mathbb{C}) / \mathcal{R}_n(\mathbb{C})$$

siendo $\mathcal{R}_n(\mathbb{C})$ el conjunto de las matrices triangulares superiores invertibles y la relación de equivalencia está dada por $X \sim Y$ sii $\exists R \in \mathcal{R}_n(\mathbb{C}) / XR = Y$. Denotaremos por $\langle X \rangle$ a la clase de equivalencia de X .

Digresión $\text{Flag}(\mathbb{C}^n)$ es un espacio homogéneo dado que $\mathcal{R}_n(\mathbb{C})$ actúa sobre $\mathbb{G}l(n, \mathbb{C})$ multiplicando a derecha, y luego tomamos el cociente por esta acción, i.e. el espacio de órbitas.

Comentario 2.24. La identificación anterior se puede hacer partiendo del grupo de las matrices unitarias $\mathcal{U}(n) = \{U \in \mathbb{C}^{n \times n} : U^*U = \text{Id}\}$. En tal caso el cociente se realiza por $\mathcal{U}^1(n)$ de matrices unitarias diagonales. De esta manera se ve que $\text{Flag}(\mathbb{C}^n)$ es compacto.

Una propiedad muy útil para estudiar la convergencia de banderas es la siguiente.

Proposición 2.25. Sean $\{F_k\} \subset \text{Flag}(\mathbb{C}^n)$ donde $F_k = \langle U_k \rangle$, con $U_k \in \mathcal{U}(n)$ y $F = \langle U \rangle$, $U \in \mathcal{U}(n)$. Entonces $F_k \rightarrow F$ si y sólo si existe $\{T_k\} \subset \mathcal{U}^1(n)$ tales que $U_k T_k \rightarrow U$. \square

La prueba se deja a cargo del lector. Observar que existen espacios cocientes donde la propiedad anterior análoga, no cierta.

Definición 2.26 (Acción de $\mathbb{G}l(n, \mathbb{C})$ en espacio bandera). Dada $A \in \mathbb{G}l(n, \mathbb{C})$, definimos el mapa inducido $A_\# : \text{Flag}(\mathbb{C}^n) \rightarrow \text{Flag}(\mathbb{C}^n)$ como

$$A_\# F = (AF^{(0)}, \dots, AF^{(n)}).$$

Esto define una acción de $\mathbb{G}l(n, \mathbb{C})$ en $\text{Flag}(\mathbb{C}^n)$

Observación 2.27. Sea $F = \langle X \rangle$ la bandera generada por la matriz X entonces $A_\# F = \langle AX \rangle$.

Proposición 2.28. Sean $A \in \mathbb{G}l(n, \mathbb{C})$ y $F = \langle Q \rangle \in \text{Flag}(\mathbb{C}^n)$ con $Q \in \mathcal{U}(n)$. Entonces F es punto fijo de $A_\#$ si y sólo si $\exists R \in \mathcal{R}_n(\mathbb{C})$ tal que $A = QRQ^*$.

Demostración. Por la observación anterior, deducimos que Q es punto fijo de $A_\#$ si y sólo si $\langle Q \rangle = \langle AQ \rangle$, equivalentemente, si y sólo si existe $R \in \mathcal{R}_n(\mathbb{C})$ tal que $AQ = QR$. Es decir, si y sólo si existe $R \in \mathcal{R}_n(\mathbb{C})$ tal que $A = QRQ^*$. \square

De la Proposición 2.28 resulta que $\langle Q \rangle \in \text{Flag}(\mathbb{C}^n)$, con $Q \in \mathcal{U}(n)$, es punto fijo de $A_\#$ si y sólo si $\exists R \in \mathcal{R}_n(\mathbb{C})$ tal que $A = QRQ^*$. (Descomposición de Schur de la matriz A .)

El teorema principal de esta sección es el siguiente.

Teorema 2.29 (Shub-Vasquez). *Supongamos $A \in \mathbb{G}l(n, \mathbb{C})$ es diagonalizable con valores propios de módulo distinto. Entonces se tiene*

1. $A_\# : \text{Flag}(\mathbb{C}^n) \rightarrow \text{Flag}(\mathbb{C}^n)$ tiene $n!$ puntos fijos.
2. $\forall F \in \text{Flag}(\mathbb{C}^n)$, $A_\#^n(F)$ es convergente.
3. La cuenca de atracción de uno de los puntos fijos es abierta y densa.

Demostración. Primero veamos que podemos restringirnos al caso diagonal.

Como A es diagonalizable, existe $X \in \mathbb{G}l(n, \mathbb{C})$ tal que $A = XDX^{-1}$. Luego $A_\# = X_\#D_\#(X^{-1})_\#$, donde $(X^{-1})_\# = (X_\#)^{-1}$. Como el mapa $X_\# : \text{Flag}(\mathbb{C}^n) \rightarrow \text{Flag}(\mathbb{C}^n)$ es un homeo, resulta que $A_\#$ y $D_\#$ son topológicamente conjugados, por tanto tienen misma dinámica.

Ordenamos $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ para que $|\lambda_1| > \dots > |\lambda_n|$.

Sea $\mathcal{P} \subset \mathbb{G}l(n, \mathbb{C})$ el grupo discreto de permutaciones. Observar que $\#\mathcal{P} = n!$. Además, si $\langle P \rangle = \langle P' \rangle$ entonces $P = P'T$ para cierta $T \in \mathcal{U}^1(n)$, por lo tanto $P^{-1}P' = T$. Al ser \mathcal{P} un grupo y $\mathcal{P} \cap \mathcal{U}(n) = \{\text{Id}\}$ se tiene que $P = P'$. Es decir, $\#\langle \mathcal{P} \rangle = n!$.

Observar que $D_\#\langle P \rangle = \langle DP \rangle = \langle P \rangle$. Ya que por ser D diagonal y $P \in \mathcal{P}$, existe D' diagonal, tal que $DP = PD'$. Por lo tanto $\langle P \rangle$ es punto fijo de $D_\#$ para toda $P \in \mathcal{P}$.

Sea $W^s(\langle P \rangle)$ la variedad estable asociada al punto fijo $\langle P \rangle$, es decir

$$W^s(\langle P \rangle) = \{ \langle X \rangle \mid \lim_{k \rightarrow +\infty} D_\#^k \langle X \rangle = \langle P \rangle \}.$$

Sean $\mathcal{L}_n(\mathbb{C})$ el conjunto de las matrices triangulares inferiores invertibles. Dada $L \in \mathcal{L}_n(\mathbb{C})$ tenemos

$$D_\#^k \langle LP \rangle = \langle D^k LP \rangle = \langle D^k LD^{-k} D^k P \rangle = \langle D^k LD^{-k} PD' \rangle = \langle D^k LD^{-k} P \rangle.$$

Ahora, si $L = (l_{ij})$, entonces $(D^k LD^{-k})_{ij} = (\frac{\lambda_i}{\lambda_j})^k l_{ij}$. Luego, como $|\lambda_1| > \dots > |\lambda_n|$ se tiene $D^k LD^{-k} \rightarrow \text{diag}(l_{11}, \dots, l_{nn}) = \tilde{L}$, entonces $D^k LD^{-k} P \rightarrow \tilde{L}P$. Pasando al cociente concluimos $D_\#^k \langle LP \rangle \rightarrow \langle \tilde{L}P \rangle = \langle P \rangle$. Es decir $\mathcal{L}_n(\mathbb{C}) \langle P \rangle = \{ \langle LP \rangle : L \in \mathcal{L}_n(\mathbb{C}) \} \subset$

$W^s(\langle P \rangle)$.

Ahora aplicando la eliminación gaussiana con pivot, dada $X \in \text{Gl}(n, \mathbb{C})$, existen $L \in \mathcal{L}_n(\mathbb{C})$, $P \in \mathcal{P}$, $R \in \mathcal{R}_n(\mathbb{C})$ tales que $X = LPR$, por lo tanto $\langle X \rangle = \langle LP \rangle$. Se concluye que

$$\text{Flag}(\mathbb{C}^n) = \bigcup_{p \in \mathcal{P}} W^s(\langle P \rangle).$$

De esto se deducen las primeras dos afirmaciones. La tercera, se desprende de observar con cuidado la eliminación gaussiana. Por ejemplo, para no tener que usar pivot es necesario que ciertas ecuaciones algebraicas no sea anulen, lo cual es una condición abierta. \square

2.2. Cociente de Rayleigh

En esta sección estudiaremos el problema de valores propios como un problema de optimización.

Sea $A \in \mathbb{R}^{n \times n}$ simétrica con valores propios $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ y vectores propios asociados v_1, \dots, v_n . El *cociente de Rayleigh* de A es la función $r_A : \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}$ dada por

$$r_A(x) = \frac{x^T A x}{x^T x} = \frac{\langle Ax, x \rangle}{\langle x, x \rangle}.$$

Observar que r_A es invariante por cambios de escala, i.e. $r_A(tx) = r_A(x)$, $\forall t \in \mathbb{R} \setminus \{0\}$. En este sentido r_A está definido también en $S^{n-1} = \{x \in \mathbb{R}^n : \|x\| = 1\}$ o $\mathbb{R}P^{n-1}$.

De ahora en adelante trabajaremos sobre S^{n-1} .

Teorema 2.30 (Courant-Fisher). *Se tiene $\lambda_1 = \max_{\|x\|=1} r_A(x)$, $\lambda_n = \min_{\|x\|=1} r_A(x)$. Además, los puntos críticos de r_A en S^{n-1} son los vectores propios v_i de A .*

Demostración. Como r_A es continua, de hecho es \mathcal{C}^∞ , y S^{n-1} compacto, tenemos máximo y mínimo.

Tenemos que el diferencial $Dr_A(x) : T_x S^{n-1} \rightarrow \mathbb{R}$, donde $r_A(x) = \langle Ax, x \rangle$, está dado por $Dr_A(x)\dot{x} = 2\langle Ax, \dot{x} \rangle$. (Acá estamos haciendo derivada libre de r_A como si estuviera viviendo en \mathbb{R}^n .)

Si denotamos por $\nabla r_A(x)$ el gradiente libre, tenemos $\nabla r_A(x) = Ax$.

Si $\text{grad} r_A(x)$ es el gradiente de r_A definido por la métrica inducida en la esfera, resulta que $\text{grad} r_A(x)$ es la proyección ortogonal de la derivada libre en el subespacio ortogonal da x . Esto es

$$\text{grad} r_A(x) = (\text{Id}_n - xx^T)Ax = Ax - \langle Ax, x \rangle x.$$

Entonces tenemos que $x \in S^{n-1}$ es punto crítico de r_A si y sólo si $\text{grad} r_A(x) = 0$. Es decir, x es punto crítico de r_A si y sólo si x es vector propio con valor propio $\langle Ax, x \rangle = r_A(x)$.

Luego el resultado sigue de que el máximo y el mínimo se alcanzan en puntos críticos. \square

Comentario 2.31. El Teorema 2.30 también se puede obtener utilizando multiplicadores de Lagrange. Sea $g : \mathbb{R}^n \rightarrow \mathbb{R}$, $g(x) = \|x\|^2 - 1$, se tiene $S^{n-1} = g^{-1}(0)$. Los puntos críticos restringidos a S^{n-1} se encuentran cuando $\lambda \nabla g(x) = \nabla r_A(x)$, es decir, $\lambda x = Ax$.

2.2.1. Flujo gradiente de Rayleigh

Tenemos que S^{n-1} es una variedad Riemanniana con la métrica inducida por el ambiente, luego podemos definir el flujo gradiente asociado a r_A sobre S^{n-1} por

$$\dot{x} = \text{grad} r_A(x) = (A - r_A(x)\text{Id}_n)x. \quad (2.7)$$

(Hemos elegido a propósito el signo positivo del gradiente.)

Observar que si $x(t)$ es una solución de (2.7), entonces está definida para todo tiempo (ver Teorema B.52). Además si $x(0) \in S^{n-1}$, entonces $x(t) \in S^{n-1}$ para todo $t \in \mathbb{R}$. En efecto,

$$\frac{1}{2} \frac{d}{dt} \|x(t)\|^2 = \langle \dot{x}(t), x(t) \rangle = \langle \text{grad} r_A(x(t)), x(t) \rangle = 0,$$

dado que $\text{grad} r_A(x(t)) \in T_{x(t)} S^{n-1} = x(t)^\perp$.

El objetivo de esta sección es estudiar la convergencia de este flujo.

Teorema 2.32. *El flujo gradiente (+) de r_A en S^{n-1} con respecto a la métrica inducida por la euclídea está dado por $\dot{x} = (A - r_A(x))\text{Id}_n x$. Sus soluciones están definidas para todo $t \in \mathbb{R}$ y convergen a un vector propio de A . Si $\lambda_1 > \lambda_2$, entonces para “casi toda” condición inicial x_0 , $\|x_0\| = 1$, $x(t)$ converge a v_1 (o $-v_1$).*

Demostración. Veamos que las soluciones del flujo son proyecciones en la esfera de soluciones de $\dot{x} = Ax$.

Sea $\gamma(t)$ solución de $\dot{x} = Ax$ en \mathbb{R}^n , con condición inicial $\gamma(0) \in S^{n-1}$ y consideremos la curva $\eta(t) = \gamma(t)/\|\gamma(t)\|$. Entonces derivando tenemos

$$\dot{\eta} = \frac{\dot{\gamma}}{\|\gamma\|} + \frac{d}{dt} \left(\frac{1}{\|\gamma(t)\|} \right) \gamma = \frac{\dot{\gamma}}{\|\gamma\|} - \frac{\langle \gamma, \dot{\gamma} \rangle}{\|\gamma\|^2} \frac{\gamma}{\|\gamma\|}$$

usando que $\dot{\gamma} = A\gamma$ tenemos

$$\dot{\eta} = A \frac{\gamma}{\|\gamma\|} - \frac{\langle \gamma, A\gamma \rangle}{\|\gamma\|^2} \frac{\gamma}{\|\gamma\|} = (A - r_A(\gamma)) \frac{\gamma}{\|\gamma\|}$$

y dado que r_A es invariante por escalas tenemos

$$\dot{\eta} = (A - r_A(\eta))\eta.$$

Por lo tanto, concluimos que $\eta(t) = e^{At}x_0/\|e^{At}x_0\|$ es la solución del flujo gradiente con condición inicial $\eta(0) = \frac{x_0}{\|x_0\|} \in S^{n-1}$. Observar que en $\mathbb{R}P^{n-1}$ la solución análoga es $e^{At}x_0$.

Al igual que antes, basta suponer para estudiar la dinámica que $A = \text{diag}(\lambda_1, \dots, \lambda_n)$. Es un ejercicio ver que la solución así obtenida verifica las condiciones del enunciado (similar al caso del método de la potencia). \square

◇ **2.33.** Probar que los puntos críticos v_i , ($i = 2, \dots, n-1$), son puntos silla.

Sugerencia: acercarse por dos curvas distintas donde la f sea mínimo y máximo local respectivamente. O calcular directamente el Hessiano (ver Sección D.7).

◇ **2.34.** Probar que r_A es una función de Morse. Observar que de esta manera podemos aplicar el Teorema B.52.

Uno se podría preguntar si es posible extender el flujo de Rayleigh a $\mathbb{R}^n \setminus \{0\}$. La respuesta es afirmativa. Una forma posible es cambiar la métrica Riemanniana

de $\mathbb{R}^n \setminus \{0\}$.

Dado $x \in \mathbb{R}^n \setminus \{0\}$, consideramos el producto interno $\langle \cdot, \cdot \rangle_x$ en $T_x(\mathbb{R}^n \setminus \{0\})$ como

$$\langle \zeta, \eta \rangle_x := \frac{\langle \zeta, \eta \rangle}{\|x\|^2}, \quad \zeta, \eta \in T_x(\mathbb{R}^n \setminus \{0\})$$

siendo $\langle \cdot, \cdot \rangle$ la métrica Euclideana en \mathbb{R}^n . Es fácil verificar que $\langle \cdot, \cdot \rangle_x$ define una métrica Riemanniana.

Dado que un producto interno sobre un espacio vectorial esta en correspondencia con formas cuadráticas ($\langle \zeta, \eta \rangle = \eta^T Q \zeta$ con Q simétrica definida positiva), existe $Q(x)$ de modo que $\langle \zeta, \eta \rangle_x = \eta^T Q(x) \zeta$. En nuestro caso $Q(x) = \|x\|^{-2} \text{Id}_n$.

Para $r_A(x) = \langle Ax, x \rangle / \|x\|^2$, $x \in \mathbb{R}^n \setminus \{0\}$, usando el Lema B.48 resulta $\text{grad } r_A(x) = \|x\|^2 \nabla r_A(x)$. Luego si calculamos el gradiente (Euclideano) $\nabla r_A(x)$ de $r_A(x)$ resulta

$$\text{grad } r_A(x) = 2(Ax - r_A(x)). \quad (2.8)$$

Comparando (2.8) con (2.7), concluimos que el flujo de Rayleigh en S^{n-1} puede extenderse a todo \mathbb{R}^n .

2.3. Cociente de Rayleigh generalizado

Dada A una matriz diagonalizable con valores propios ordenados $\lambda_1 \geq \dots \geq \lambda_n$, el subespacio k -dimensional generado por los vectores propios asociados a $\lambda_1, \dots, \lambda_k$ se denomina *subespacio dominante de dimensión k*

Para simplificar nuestro problema consideramos $A \in \mathbb{R}^{n \times n}$ simétrica. Sean $\lambda_1 \geq \dots \geq \lambda_n$ sus valores propios.

Un espacio natural para estudiar este problema es considerar la *variedad de Stiefel*

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

Observar que $X \in St(k, n)$ si las columnas son un conjunto ortonormal.

Definición 2.35 (Cociente generalizado de Rayleigh). Una función “costo” que se maximice en el subespacio dominante de dimensión k es el *cociente generalizado*

de *Rayleigh*, a saber,

$$R_A : St(k, n) \rightarrow \mathbb{R}, \quad R_A(X) = \text{tr}(X^T A X),$$

donde $\text{tr}()$ es la función traza.

Definición 2.36 (Producto interno de Frobenius). El *producto interno de Frobenius* en $\mathbb{R}^{m \times k}$ está definido por $\langle A, B \rangle_F = \text{tr}(A^T B)$.

◇ **2.37.** Es fácil ver que para matrices $A = ((a_{ij}))$, $B = ((b_{ij}))$ se tiene $\langle A, B \rangle_F := \sum_{\substack{i=1, \dots, m \\ j=1, \dots, l}} a_{ij} b_{ij}$. Esto es, coincide con el inducido por la identificación $\mathbb{R}^{m \times l} \cong \mathbb{R}^{ml}$

De esta manera resulta que el cociente generalizado de Rayleigh está dado por

$$R_A(X) = \langle AX, X \rangle_F.$$

El objetivo de esta sección es estudiar esta función y tomar el flujo gradiente asociado.

Para tal fin comenzaremos estudiando la geometría de la variedad de Stiefel.

2.3.1. Geometría de la variedad de Stiefel

La *variedad de Stiefel* es el conjunto definido por

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

Observar que $X \in St(k, n)$ si las columnas son un conjunto ortonormal.

Comencemos esta subsección estudiando la geometría de la variedad de Stiefel.

Observar que cuando $k = 1$ o n , obtenemos que $St(1, n) = S^{n-1}$ y $St(n, n) = \mathcal{O}(n)$ siendo $\mathcal{O}(n)$ el grupo ortogonal.

Comentario 2.38. Así como está definida, la variedad de Stiefel esta "encajada" en $\mathbb{R}^{n \times k}$. Sin embargo es posible obtenerla como una variedad cociente del grupo ortogonal de la siguiente manera (ver Sección C.4 de espacios homogéneos, Ejemplo C.39-6). Identificando dos matrices ortogonales por tener las primeras

k columnas iguales, o equivalentemente, si están relacionadas por multiplicar a derecha por una matriz de la forma $\begin{pmatrix} \text{Id}_k & 0 \\ 0 & Q \end{pmatrix}$ con $Q \in \mathcal{O}(n-k)$. En este sentido $St(k, n) \cong \mathcal{O}(n)/\mathcal{O}(n-k)$.

Comentario 2.39. Observar que con esta identificación de $St(k, n)$ podemos obtener una representación de la grassmaniana $Gr(k, n)$ identificando $X \in St(k, n)$ con XQ , donde $Q \in \mathcal{O}(n-k)$, dado que el espacio columnas genera el mismo subespacio. Entonces obtenemos $Gr(k, n) \cong \mathcal{O}(n)/\mathcal{O}(k) \times \mathcal{O}(n-k)$.

Proposición 2.40. *La variedad de Stiefel $St(k, n)$ es una variedad diferenciable compacta, de dimensión $kn - k(k+1)/2$. Además, el espacio tangente a $X \in St(k, n)$ está dado por*

$$T_X St(k, n) = \{ \zeta \in \mathbb{R}^{n \times k} : \zeta^T X + X^T \zeta = 0 \}.$$

Demostración. Veamos que podemos obtener $St(k, n)$ como la preimagen de un valor regular de un mapa. Sea $F : \mathbb{R}^{n \times k} \rightarrow \mathbb{R}^{\frac{k(k+1)}{2}} \equiv \text{Sym}(k)$ dado por $F(X) = X^T X - \text{Id}_k$, donde $\text{Sym}(k)$ es el subespacio lineal de $\mathbb{R}^{k \times k}$ de las matrices simétricas, al cual lo identificamos de manera natural con $\mathbb{R}^{\frac{k(k+1)}{2}}$.

Afirmación: 0 es valor regular de F .

La derivada de F en $X \in F^{-1}(0) = St(k, n)$, en la dirección $\zeta \in \mathbb{R}^{n \times k}$ está dada por

$$DF(X)\zeta = \zeta^T X + X^T \zeta \in \mathbb{R}^{\frac{k(k+1)}{2}}, \quad \forall \zeta \in \mathbb{R}^{n \times k}.$$

Veamos que $DF(X)$ es sobre. Si $\eta \in \text{Im } DF(X)^\perp \subset \mathbb{R}^{\frac{k(k+1)}{2}}$, entonces

$$\langle \eta, \zeta^T X + X^T \zeta \rangle_F = 0, \quad \forall \zeta \in \mathbb{R}^{n \times k}.$$

Es decir,

$$0 = \text{tr}(\eta \zeta^T X) + \text{tr}(\eta X^T \zeta), \quad \forall \zeta \in \mathbb{R}^{n \times k}. \quad (2.9)$$

Utilizando la propiedad $\text{tr}(AB) = \text{tr}(BA)$ (siempre que el producto tenga sentido), resulta

$$\text{tr}(\eta X^T \zeta) = \text{tr}(X^T \zeta \eta) = \text{tr}((\zeta^T X)^T \eta) = \text{tr}(\eta \zeta^T X).$$

Esto implica que $\eta X^T = 0$. Multiplicando a derecha por X se obtiene $\eta = \eta X^T X = 0$. Luego 0 es valor regular.

De la afirmación se concluye que $St(k, n)$ es una variedad diferenciable de dimensión $kn - k(k+1)/2$. La compacidad resulta de ser cerrado y acotado en $\mathbb{R}^{n \times k}$.

Además el tangente $T_X St(k, n)$ está dado por $\text{Ker } DF(X) = \{\zeta \in \mathbb{R}^{n \times k} : \zeta^T X + X^T \zeta = 0\}$. \square

Observación 2.41. El tangente $T_X St(k, n)$ puede ser obtenido derivando curvas sobre $St(k, n)$. Esto es, si $\gamma(t) \in St(k, n)$, con $\gamma(0) = X$ y $\dot{\gamma}(0) = \zeta$, diferenciando $\gamma(t)^T \gamma(t) = \text{Id}_k$ obtenemos

$$0 = \left. \frac{d}{dt} \right|_{t=0} \gamma(t)^T \gamma(t) = \dot{\gamma}(0)^T \gamma(0) + \gamma(0)^T \dot{\gamma}(0) = \zeta^T X + X^T \zeta.$$

Observación 2.42. $\zeta \in T_X St(k, n)$ sii $\zeta^T X$ es antisimétrica ($k \times k$)

Definición 2.43 (Métrica en variedad de Stiefel). Dado que $St(k, n)$ es una subvariedad encajada en $\mathbb{R}^{n \times k}$, podemos inducir la métrica Riemanniana del ambiente, a saber: dados $\zeta, \eta \in T_X St(k, n)$ definimos $\langle \zeta, \eta \rangle_X := \text{tr}(\zeta^T \eta)$.

Espacio normal $T_X St(k, n)^\perp$

Para poder calcular gradientes respecto a la métrica inducida es necesario identificar el espacio normal a $T_X St(k, n)$. Definimos el espacio normal

$$T_X St(k, n)^\perp := \{N \in \mathbb{R}^{n \times k} : \text{tr}(N^T \zeta) = 0, \forall \zeta \in T_X St(k, n)\}.$$

Es claro que su dimensión es $k(k+1)/2$.

Lema 2.44. $T_X St(k, n)^\perp = \{X\Lambda \in \mathbb{R}^{n \times k} : \Lambda^T = \Lambda \in \mathbb{R}^{k \times k}\}$

Demostración. Observar que

$$\text{tr}((X\Lambda)^T \zeta) = \text{tr}(\Lambda X^T \zeta) = \text{tr}(X^T \zeta \Lambda) = \text{tr}((\zeta^T X)^T \Lambda) = \text{tr}(\Lambda \zeta^T X)$$

entonces

$$\text{tr}((X\Lambda)^T \zeta) = \frac{1}{2}(\text{tr}(\Lambda X^T \zeta + \Lambda \zeta^T X)) = \frac{1}{2} \text{tr}(\Lambda(X^T \zeta + \zeta^T X)) = 0$$

pues $\zeta \in T_X St(k, n)$. \square

Recordar que si tenemos $A \in \mathbb{R}^{k \times k}$ se define la *parte simétrica* $\text{Sym}(A) = \frac{1}{2}(A + A^T)$ y la *parte antisimétrica* como $\text{Skew}(A) = \frac{1}{2}(A - A^T)$.

Proposición 2.45. Sea $\pi_X : \mathbb{R}^{n \times k} \rightarrow T_X St(k, n)$ la proyección ortogonal en el tangente. Entonces, si $Z \in \mathbb{R}^{n \times k}$, se tiene

$$\pi_X(Z) = X \text{Skew}(X^T Z) + (I - XX^T)Z.$$

Además $\pi_X^\perp(Z) = X \text{Sym}(X^T Z)$, siendo π_X^\perp la proyección ortogonal en $T_X St(k, n)^\perp$.

Demostración. Veamos primero que la proyección en $T_X St(k, n)^\perp$ es $X \text{Sym}(X^T Z)$. Por el lema anterior es claro que $X \text{Sym}(X^T Z) \in T_X St(k, n)^\perp$. Resta ver que $Z - X \text{Sym}(X^T Z) \in T_X St(k, n)$ y que la fórmula es correcta.

Se tiene que

$$\begin{aligned} X^T(Z - X \text{Sym}(X^T Z)) + (Z - X \text{Sym}(X^T Z))^T X \\ = X^T Z - X^T X \text{Sym}(X^T Z) + Z^T X - \text{Sym}(X^T Z) X^T X \\ = X^T Z - 2 \text{Sym}(X^T Z) + Z^T X = 0. \end{aligned}$$

Además se tiene

$$\begin{aligned} \pi_X(Z) &:= X \text{Skew}(X^T Z) + (I - XX^T)Z \\ &= X \left(\frac{X^T Z - Z^T X}{2} \right) + Z - XX^T Z \\ &= X \left(\frac{-X^T Z - Z^T X}{2} \right) + Z = Z - \pi_X^\perp(Z). \end{aligned}$$

□

Observación 2.46. De la expresión de la proyección surge lo siguiente. Sea X_\perp una matriz $n \times (n-k)$ que complete X para que $[X, X_\perp] \in \mathcal{O}(n)$. Entonces se puede ver que toda matriz $Z \in \mathbb{R}^{n \times k}$ se puede escribir como

$$Z = XA + X_\perp B, \quad A \in \mathbb{R}^{k \times k}, \quad B \in \mathbb{R}^{(n-k) \times k}.$$

Luego, se puede ver lo siguiente

$$Z = XA + X_\perp B \in T_X St(k, n) \iff A^T = -A.$$

De esta manera el subespacio $T_X St(k, n)$ puede parametrizarse con $A \in \text{Antisim} \equiv \mathbb{R}^{\frac{k(k+1)}{2}}$ y $B \in \mathbb{R}^{(n-k) \times k}$. En particular, cuando $X = \begin{pmatrix} \text{Id}_k \\ 0 \end{pmatrix}$, el espacio tangente se identifica con $\zeta = \begin{pmatrix} A \\ B \end{pmatrix}$, con $A \in \mathbb{R}^{k \times k}$ antisimétrica y $B \in \mathbb{R}^{(n-k) \times k}$ cualquiera.

2.3.2. Aspectos geométricos y dinámicos del flujo

Volvamos a trabajar con nuestro problema original.

Teorema 2.47 (Flujo generalizado de Rayleigh). *Sea $A \in \mathbb{R}^{n \times n}$, simétrica, $\lambda_1 \geq \dots \geq \lambda_n$ los valores propios con v_1, \dots, v_n vectores propios respectivos. Entonces el cociente generalizado de Rayleigh R_A tiene $\binom{n}{k}$ puntos críticos de la forma $X_I = (v_{i_1}, \dots, v_{i_k})$ con $1 \leq i_1 < \dots < i_k \leq n$. Además el resto de los puntos críticos son de la forma $UX_I V$ con $U \in \mathcal{O}(n)$, $V \in \mathcal{O}(k)$ tales que $U^T A U = A$. El valor crítico de R_A asociado a X_I es $\lambda_{i_1} + \dots + \lambda_{i_k}$. En particular $\max R_A = \lambda_1 + \dots + \lambda_k$ y $\min R_A = \lambda_{n-k+1} + \dots + \lambda_n$.*

Demostración. Calculemos $\text{grad} R_A$. Para eso podemos extender R_A a $\mathbb{R}^{n \times k}$ con la misma expresión. Sea \tilde{R}_A dicha extensión. Sea $\nabla \tilde{R}_A$ el gradiente libre de la extensión de R_A . Luego proyectando en el espacio tangente $T_X St(k, n)$, resulta, $\text{grad} R_A(X) = \pi_X \nabla \tilde{R}_A(X)$, con $X \in St(k, n)$, donde π_X está dada en Proposición 2.45. Es fácil ver que $D\tilde{R}_A(X)\dot{X} = 2\langle AX, \dot{X} \rangle$ por lo que $\nabla \tilde{R}_A(X) = 2AX$. Luego usando Proposición 2.45 se tiene

$$\frac{1}{2} \text{grad} R_A(X) = \pi_X(AX) = X \text{Skew}(X^T AX) + (\text{Id}_n - XX^T)AX = (\text{Id}_n - XX^T)AX,$$

ya que $X^T AX$ es simétrica.

Observar que la transformación $\text{Id}_n - XX^T$ vista como transformación lineal de \mathbb{R}^n en \mathbb{R}^n proyecta los vectores en X^\perp , el subespacio ortogonal al generado por las columnas de X . Luego $\text{grad} R_A(X) = 0$ sii el espacio de columnas de AX es un subespacio del espacio de columnas de X , i.e. si el espacio de columnas de X es A -invariante. Lo que resta de la prueba es la caracterización de las X de tal forma que sus espacios de columnas sean A -invariantes, lo cual se deja de ejercicio. \square

\diamond **2.48.** En el contexto del teorema anterior, probar que el espacio de columnas de X es A -invariante si y solo si X es de la forma $UX_I V$ con $U \in \mathcal{O}(n)$, $V \in \mathcal{O}(k)$ tales que $U^T A U = A$

Flujo gradiente asociado

Teorema 2.49 (Flujo gradiente asociado). *El flujo gradiente de $R_A : St(k, n) \rightarrow \mathbb{R}$ está dado por*

$$\dot{X} = (\text{Id}_n - XX^T)AX.$$

Las soluciones existen para todo $t \in \mathbb{R}$ y convergen a subespacios propios generalizados $(UX_I V)$ como antes.

Si $\lambda_k > \lambda_{k+1}$ entonces R_A es una función de Morse-Bott. En particular para casi toda $X_0 \in St(k, n)$, $X(t)$ curva de flujo con dato inicial X_0 , converge a $UX_{(1, \dots, k)}V$. Su velocidad es exponencial $\lambda_k - \lambda_{k+1}$, localmente cerca de subvariedad crítica.

Demostración. Este teorema se deduce del Comentario B.54 en página 130 sobre flujos gradientes y funciones de Morse-Bott. \square

2.4. Algoritmo QR

El algoritmos *QR* es el más simple y conocido para encontrar valores propios de una matriz.

Este algoritmo está basado en la descomposición QR de una matriz.

Definición 2.50 (Descomposición QR). Sea $A \in \text{Gl}(n, \mathbb{C})$ (o $\text{Gl}(n, \mathbb{R})$). La *descomposición QR* de A es la descomposición $A = QR$, donde Q es una matriz unitaria (o ortogonal) y R una triangular superior.

La existencia de tal descomposición resulta del método de Gram-Schmidt. Sin embargo la descomposición no es única a menos que se pida que las entradas diagonales de R sean positivas.

◇ **2.51.** Probar que toda matriz tiene una descomposición QR. Además, si tenemos dos descomposiciones $A = QR = Q'R'$, entonces $Q' = QT^*$ y $R' = TR$, para T diagonal unitaria (o diagonal ortogonal en el caso real). Si los elementos de la diagonal de R son positivos, entonces la descomposición es única.

Para simplificar la exposición nos restringiremos en esta sección al estudio del caso complejo. El caso real es similar.

ALGORITMO QR

- **Input:** $A \in \mathbb{G}l(n, \mathbb{C})$
- **Output:** $\{A_k\}$ en $\mathbb{G}l(n, \mathbb{C})$

>> $A_0 = Q_0 R_0$ (descomposición QR)

>> **for** $k = 0, 1, 2, \dots$

$$A_{k+1} =: R_k Q_k = Q_{k+1} R_{k+1}$$

Observar que $A_1 = R_0 Q_0 = Q_0^T Q_0 R_0 Q_0 = Q_0^T A Q_0$, y por lo tanto el espectro de A_1 coincide con el espectro de A_0 . De manera análoga se observa que $A_2 = (Q_0 Q_1)^T A_0 (Q_0 Q_1)$, y por lo tanto A_2 tiene mismo espectro que A_0 . Procediendo de esta manera, el algoritmo QR produce una sucesión $\{A_k\}$ tal que,

$$A_{k+1} = \tilde{Q}_k^* A_0 \tilde{Q}_k, \quad \text{donde} \quad \tilde{Q}_k = Q_0 \cdots Q_k \in \mathcal{U}(n) \quad (2.10)$$

En resumen, el algoritmo QR produce una sucesión $\{A_k\}$ de matrices, unitariamente conjugadas por \hat{Q}_k , y por tanto con mismo espectro que $A = A_0$.

Lema 2.52. Dado $k \in \mathbb{N}$, resulta $A^{k+1} = \tilde{Q}_k \tilde{R}_k$, para cierta matriz triangular \tilde{R}_k .

Demostración. Observar primero que

$$A^{k+1} = (Q_0 R_0)^{k+1} = Q_0 (R_0 Q_0)^k R_0 = Q_0 A_1^k R_0 = Q_0 (Q_1 R_1)^k R_0 = Q_0 Q_1 (R_1 Q_1^{k-1}) R_1 R_0,$$

procediendo de esta manera se obtiene

$$A^{k+1} = Q_0 \cdots Q_{k-1} (R_{k-1} Q_{k-1}) R_{k-1} \cdots R_0 = Q_0 \cdots Q_k R_k \cdots R_0 = \tilde{Q}_k \tilde{R}_k.$$

□

Observación 2.53. Lo que el lema anterior nos dice es que \tilde{Q}_k es la componente unitaria de la descomposición QR de A^{k+1} . Además, como una matriz triangular preserva en particular la bandera de la identidad ($\langle e_1 \rangle, \langle e_1, e_2 \rangle, \dots, \langle e_1, \dots, e_n \rangle$) resulta lo siguiente.

Lema 2.54. Se tiene $A_{\#} \langle \tilde{Q}_k \rangle = \langle \tilde{Q}_{k+1} \rangle$

Demostración.

$$A_{\#}\langle \tilde{Q}_k \rangle = \langle A\tilde{Q}_k \rangle = \langle A\tilde{Q}_k\tilde{R}_k \rangle = \langle AA^{k+1} \rangle = \langle \tilde{Q}_{k+1}\tilde{R}_{k+1} \rangle = \langle \tilde{Q}_{k+1} \rangle$$

□

Teorema 2.55 (Dinámica QR). *Sea $A \in \mathbb{G}l(n, \mathbb{C})$ con valores propios de módulo distinto. La sucesión $\{A_k\}$ dada en 2.10 por el algoritmo QR satisface*

- $\text{diag}A_k$ tiende a una matriz diagonal de valores propios de A .
- $(A_k)_{ij} \rightarrow 0$ si $i > j$;
- los módulos de $(A_k)_{ij}$ son convergentes para $i < j$.

Demostración. Del Lema 2.54 se tiene que $A_{\#}^{k+1}\langle \text{Id} \rangle = \langle \tilde{Q}_k \rangle$. Luego, por el Teorema 2.29 de Shub-Vasquez resulta

$$\langle \tilde{Q}_k \rangle \rightarrow \langle \tilde{Q} \rangle \in \text{Flag}(\mathbb{C}^n)$$

siendo $\langle \tilde{Q} \rangle$ punto fijo de $A_{\#}$. Esto es $\tilde{Q}^*A\tilde{Q} = \tilde{R}$ con \tilde{R} triangular superior. Luego por como es la convergencia en $\text{Flag}(\mathbb{C}^n)$, existen $T_k \in \mathcal{U}^1(n)$ tales que

$$\tilde{Q}_k T_k \rightarrow \tilde{Q} \text{ en } \mathcal{U}(n).$$

Entonces

$$T_k^* (\tilde{Q}_k^* A \tilde{Q}_k) T_k = (\tilde{Q}_k T_k)^* A (\tilde{Q}_k T_k) \xrightarrow{k} \tilde{Q}^* A \tilde{Q}.$$

Luego, si denotamos $T = \text{diag}(\theta_1^{(k)}, \dots, \theta_n^{(k)})$, con $|\theta_i^{(k)}| = 1$, resulta

$$(A_{k+1})_{ij} \left(\overline{\theta_i^{(k)}} \right)^i \theta_j^{(k)} \rightarrow \tilde{R}_{ij}.$$

El resultado sigue de discutir según i y j . □

Corolario 2.56. *Sea $A \in \text{Sym}(n)$ genérica con valores propios distintos, entonces A_k tiende a una matriz diagonal y, en particular \tilde{Q}_k tiende a la matriz de vectores propios de A .*

Comentario 2.57. El Algoritmo QR es el método más simple y estable para encontrar valores propios de matrices. La desventaja de este algoritmo es su velocidad. Por esta razón se utilizan variantes con “shift” para acelerar el proceso.

2.4.1. Algoritmo QR y flujo asociado

Nos restringimos al caso simétrico, $A \in \text{Sym}(n) = \{A \in \mathbb{R}^{n \times n} : A^T = A\}$.

Utilizaremos la representación única de la descomposición QR, $A = QR$, donde $Q \in \mathcal{O}(n)$ y R es triangular superior con entradas no negativas sobre la diagonal.

A continuación, veremos que la sucesión $\{A_k\}$ producida por el algoritmo QR puede obtenerse como los tiempos enteros de un flujo.

Definición 2.58 (Flujo isoespectral). Una ecuación diferencial matricial en $\text{Sym}(n)$, $\dot{A}(t) = f(t, A(t))$ se dice *isoespectral* si toda solución $A(t)$ es de la forma

$$A(t) = Q(t)^T A(0) Q(t),$$

con $Q(t) \in \mathcal{O}(n)$.

Observación 2.59. Si $V(t) = Q(t)^T A_0 Q(t)$, se tiene,

$$\dot{V} = \dot{Q}^T A_0 Q + Q^T A_0 \dot{Q} = \dot{Q}^T Q V + V Q^T \dot{Q},$$

además como $\dot{Q} \in T_Q \text{St}(n, n)$, tenemos que $\dot{Q}^T Q$ es antisimétrica. Denotando $\dot{Q}^T Q = B$ se tiene que

$$\dot{V} = VB - BV = [V, B],$$

siendo $[\cdot, \cdot]$ el corchete de Lie matricial. Por tanto, si $Q(t) \in \mathcal{O}(n)$ es diferenciable, definiendo $B = \dot{Q}^T Q$ resulta que V satisface la ecuación diferencial

$$\dot{V} = [V, B].$$

Proposición 2.60. Sea $B(t) \in \mathbb{R}^{n \times n}$ continua y antisimétrica. La ecuación diferencial $\dot{A}(t) = [A(t), B(t)]$ es isoespectral.

Demostración. Observar que si $A(0) = \text{Id}_n$ es la condición inicial, entonces la solución constante $A(t) = \text{Id}_n$, $t \in \mathbb{R}$, es la solución con esa condición inicial. Motivado por lo anterior, sea $Q(t)$ la solución en $\text{Sym}(n)$ de la ecuación diferencial $\dot{Q}(t) = Q(t)B(t)$. Veamos que $Q(t) \in \mathcal{O}(n)$. Consideremos la derivada de $Q(t)^T Q(t)$, se tiene

$$\begin{aligned} \frac{d}{dt} Q(t)^T Q(t) &= \dot{Q}^T Q + Q^T \dot{Q} \\ &= B^T Q^T Q + Q^T Q B \\ &= Q^T Q B - B Q^T Q = [Q^T Q, B]. \end{aligned}$$

Luego, si $Q(0)^T Q(0) = \text{Id}_n$, resulta de la observación inicial que $Q(t)^T Q(t) = \text{Id}_n$, $t \in \mathbb{R}$, y por lo tanto, $Q(t) \in \mathcal{O}(n)$.

Veamos que $\hat{A}(t) = Q(t)^T A(0) Q(t)$ es la solución de la ecuación diferencial del enunciado con condición inicial $\hat{A}(0) = A(0)$. Derivando obtenemos

$$\begin{aligned} \frac{d}{dt} \hat{A}(t) &= \dot{Q}^T A(0) Q + Q^T A(0) \dot{Q} \\ &= (QB)^T A(0) Q + Q^T A(0) (QB) \\ &= -BQ^T A(0) Q + Q^T A(0) QB \\ &= -\hat{B}\hat{A} + \hat{A}\hat{B}. \end{aligned}$$

Luego, por unicidad de las soluciones, se concluye el resultado. \square

Observación 2.61. Recordar que toda matriz $A \in \mathbb{R}^{n \times n}$ se puede factorizar de forma única como $A = A_- + A_+$ siendo A_- antisimétrica y A_+ triangular superior. Recordar también que si $A > 0$ (definida positiva) entonces tiene un logaritmo, i.e. $\log A$ es tal que $e^{\log A} = A$.

Teorema 2.62 (Flujo QR). *Sea $A_0 > 0$, la ecuación diferencial $\dot{A} = [A, (\log A)_-]$, $A(0) = A_0$ está definida $\forall t \in \mathbb{R}$. Además $A(k)$, $k \in \mathbb{N}$ coincide con A_k del algoritmo QR.*

Demostración. Observar que $A_0^k = e^{k \log A_0}$, $k \in \mathbb{N}$. Luego si consideremos la factorización QR de la curva $e^{t \log A_0} = Q(t)R(t)$, con $t \in \mathbb{R}$, resulta del Lema 2.52 que $Q(k+1) = \tilde{Q}_k$, $k \in \mathbb{N}$ (cf. (2.10)). Luego la curva $A(t) = Q(t)^T A_0 Q(t)$ satisface que $A(k+1) = \tilde{Q}_k^T A_0 \tilde{Q}_k = A_{k+1}$. Sólo resta ver que la curva $A(t)$ satisface la ecuación diferencial del enunciado. En efecto, derivando $e^{t \log A_0}$ obtenemos

$$\dot{Q}R + Q\dot{R} = \log(A_0)QR,$$

luego $Q^T \dot{Q} + \dot{R}R^{-1} = Q^T \log(A_0)Q$. Como $Q^T \dot{Q}$ es antisimétrica y $\dot{R}R^{-1}$ es triangular superior, tenemos que $(Q^T \log(A_0)Q)_- = Q^T \dot{Q}$. Luego derivando $A(t)$, tenemos $\dot{A} = \dot{Q}^T A_0 Q + Q^T A_0 \dot{Q}$ por lo tanto

$$\begin{aligned} \dot{A} &= -(\log(Q^T A_0 Q))_- Q^T A_0 Q + Q^T A_0 Q (\log(Q^T A_0 Q))_- \\ &= -(\log(Q^T A_0 Q))_- A + A (\log(Q^T A_0 Q))_-. \end{aligned}$$

\square

2.5. Descomposición en valores singulares

Dada $A \in \mathbb{R}^{m \times n}$ con $m \geq n$, existen $U \in \mathcal{O}(n)$, $V \in \mathcal{O}(m)$ tales que

$$A = V\Sigma U^T, \text{ con } \Sigma = \begin{pmatrix} \text{diag}(\sigma_1, \dots, \sigma_n) \\ 0_{(m-n) \times n} \end{pmatrix}$$

siendo los $\sigma_i \geq 0$, los valores singulares de A . A esta descomposición se le llama *descomposición en valores singulares*.

Observación 2.63. Los σ_i son las raíces cuadradas de los valores propios de $A^T A$. En efecto,

$$A^T A = U\Sigma^T \Sigma U^T,$$

donde $\Sigma^T \Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$.

Esta observación sugiere una idea de la prueba de la existencia de tal descomposición. Al ser $A^T A$ diagonalizable, existe $U \in \mathcal{O}(n)$ tal que $U^T A^T A U = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ para ciertos $\sigma_i \geq 0$. Por lo tanto si $U = (u_1, \dots, u_n)$, tenemos $A^T A u_i = \sigma_i^2 u_i$. Por lo tanto para los $\sigma_i \neq 0$ definimos $v_i := A u_i / \sigma_i$. Observar que el conjunto $\{v_i\}$ es ortonormal. En efecto,

$$\langle v_i, v_j \rangle = \left\langle \frac{A u_i}{\sigma_i}, \frac{A u_j}{\sigma_j} \right\rangle = \frac{1}{\sigma_i \sigma_j} \langle A^T A u_i, u_j \rangle = \frac{\sigma_i^2}{\sigma_i \sigma_j} \langle u_i, u_j \rangle = \frac{\sigma_i}{\sigma_j} \delta_{ij}.$$

Luego, completando $\{v_i\}$ a una base ortonormal de \mathbb{R}^m resulta $A u_i = \sigma_i v_i$, para concluir que $AU = V\Sigma$.

Observación 2.64. Si A es simétrica, se tiene $\sigma_i = |\lambda_i|$, con $\lambda_i \in Sp(A)$.

Una forma de encontrar valores singulares de A es hallar valores propios de $A^T A$. Sin embargo el condicionamiento del nuevo problema puede verse afectado.

◇ **2.65.** Definiendo $\hat{A} = \begin{pmatrix} 0 & A \\ A^T & 0 \end{pmatrix} = \hat{A}^T$, entonces existe \tilde{V} ortonormal tal que

$$\tilde{V}^T \hat{A} \tilde{V} = \begin{pmatrix} \Sigma & 0 & 0 \\ 0 & -\Sigma & 0 \\ 0 & 0 & 0 \end{pmatrix} = \text{diag}(\sigma_1, \dots, \sigma_n, -\sigma_1, \dots, -\sigma_n, 0, \dots, 0),$$

donde $\tilde{V} = \frac{1}{\sqrt{2}} \begin{pmatrix} V_1 & V_1 & V_2 \\ U & -U & 0 \end{pmatrix}$, siendo $V = (V_1 \ V_2)$ con U y V tales que $A = V\Sigma U^T$.

2.5.1. Flujos isosingulares

Definición 2.66 (Flujo isosingular). Un flujo es *isosingular* si preserva los valores singulares.

¿Cómo identificar los flujos isosingulares?

Sea $A(t) = U(t)^T \Sigma V(t)$ con $U(t)$ y $V(t)$ ortonormales, entonces

$$\dot{A} = \dot{U}^T \Sigma V + U^T \Sigma \dot{V} = \dot{U}^T U A + A V^T \dot{V},$$

observar que $\dot{U}^T U$ y $V^T \dot{V}$ son antisimétricas por ser U, V ortonormales.

Lema 2.67. Sean $C(t) \in \mathbb{R}^{n \times n}$ y $D(t) \in \mathbb{R}^{m \times m}$ continuas y antisimétricas. Entonces el flujo

$$\dot{H}(t) = H(t)C(t) + D(t)H(t)$$

es iso-singular. Recíprocamente, todo flujo iso-singular es de esta forma.

Demostración. Sean $U(t), V(t)$ solución de

$$\begin{cases} \dot{U}(t) = U(t)C(t) \\ U(0) = \text{Id}_n \end{cases}, \begin{cases} \dot{V}(t) = -V(t)D(t) \\ V(0) = \text{Id}_m \end{cases}$$

Al ser C y D antisimétricas, se tiene que U y V son ortogonales.

Sea $A(t) = V(t)^T H(0) U(t)$. Entonces $A(0) = H(0)$ y

$$\dot{A} = \dot{V}^T H(0) U + V^T H(0) \dot{U} = D(t)A(t) + A(t)C(t).$$

□

2.6. Mínimos cuadrados

Sean $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ y sea

$$\Phi: \mathbb{R}^n \rightarrow \mathbb{R}, \quad \Phi(x) = \frac{1}{2} \|Ax - b\|^2.$$

Queremos encontrar el mínimo de $\Phi(x)$. Derivando obtenemos $D\Phi(x)\dot{x} = \langle A\dot{x}, Ax - b \rangle$, entonces tenemos $D\Phi(x)\dot{x} = \langle A^T(Ax - b), \dot{x} \rangle$. En particular, viendo \mathbb{R}^n como variedad Riemanniana, resulta

$$\text{grad}\Phi(x) = \nabla\Phi(x) = A^T(Ax - b),$$

por lo tanto el flujo gradiente está dado por

$$\dot{x} = A^T(Ax - b).$$

Los puntos críticos son, $A^T(Ax - b) = 0$, lo que es equivalente a que $Ax - b \in \text{Ker}(A^T) = \text{Im}(A)^\perp$ y por lo tanto los puntos críticos son las soluciones en \mathbb{R}^n de $Ax = \Pi_{\text{Im}(A)}(b)$, siendo $\Pi_{\text{Im}(A)}$ la proyección ortogonal sobre la imagen de A .

Si A es inyectiva, entonces tiene solución única. Dado que $A^T A$ es definida positiva, tenemos que es invertible y por lo tanto

$$x_{min} = (A^T A)^{-1} A^T b =: A^\dagger b.$$

Aquí A^\dagger es la *inversa de Moore-Penrose* o *pseudoinversa*.

Observar que en este caso el Hessiano de Φ es $A^T A$ y por lo tanto x_{min} es un mínimo global atractor.

Flujos Isoespectrales

En este capítulo estudiaremos el siguiente problema:

PROBLEMA: Dada $A \in \text{Sym}(n)$, consideremos el conjunto $M(A) = \{U^T A U : U \in \mathcal{O}(n)\}$ de matrices ortogonalmente equivalentes a A . Hallar

$$\min_{H \in M(A)} \|N - H\|^2,$$

para cierta N fija arbitraria.

Para estudiar este problema es necesario tener conocimientos sobre grupos de Lie y espacios homogéneos. En el Apéndice se describen los ingredientes necesarios sobre estos temas.

Sea $Q = \text{diag}(\lambda_1 \text{Id}_{n_1}, \dots, \lambda_r \text{Id}_{n_r})$ con $\lambda_1 > \dots > \lambda_r$ y $n_1 + \dots + n_r = n$.

Proposición 3.1. El conjunto $M(Q) = \{U^T Q U : U \in \mathcal{O}(n)\}$ es un espacio homogéneo, conexo, compacto de dimensión

$$\dim M(Q) = \frac{1}{2} \left(n^2 - \sum_{i=1}^r n_i^2 \right).$$

Demostración. Consideremos la acción de $\mathcal{O}(n) \simeq \mathbb{R}^{n \times n}$ dada por $U \cdot A = U^T A U$. Esta acción es una acción de grupos de Lie (ver Definición C.24). Luego $M(Q)$

es la órbita de la acción $\mathcal{O}(n) \cdot Q$. Tomando $M(Q)$ como conjunto obtenemos que $\mathcal{O}(n) \curvearrowright M(Q)$ transitivamente. Por lo tanto $M(Q)$ es el espacio homogéneo $\mathcal{O}(n)/\mathcal{O}(n)_Q$ siendo $\mathcal{O}(n)_Q$ el subgrupo de isotropía $\{U \in \mathcal{O}(n) : U^T Q U = Q\}$.

Afirmación: Es fácil ver que $\mathcal{O}(n)_Q \cong \mathcal{O}(n_1) \times \cdots \times \mathcal{O}(n_r)$. (Probarlo.)

Además se puede verificar que $SO(n) \cdot Q = \mathcal{O}(n) \cdot Q$ por lo tanto $M(Q)$ es conexo.

Hemos concluido que $M(Q) \cong \mathcal{O}(n)/\mathcal{O}(n_1) \times \cdots \times \mathcal{O}(n_r)$ y por lo tanto $M(Q)$ es una variedad diferenciable compacta conexa de dimensión

$$\begin{aligned} n^2 - \frac{n(n+1)}{2} - \sum_{i=1}^r n_i^2 - \frac{n_i(n_i+1)}{2} &= \frac{n(n-1)}{2} - \sum_{i=1}^r \frac{n_i(n_i-1)}{2} \\ &= \frac{1}{2} (n^2 - n - \sum_{i=1}^r n_i^2 - n_i) \\ &= \frac{1}{2} (n^2 - \sum_{i=1}^r n_i^2). \end{aligned}$$

□

Comentario 3.2. La estructura diferenciable de $M(Q)$ coincide con la inducida como subvariedad encajada de $\mathbb{R}^{n \times n}$.

Lema 3.3. El espacio tangente de $M(Q)$ en H , está dado por

$$T_H M(Q) = \{[H, \Omega] : \Omega^T = -\Omega \in \mathbb{R}^{n \times n}\}.$$

Demostración. Fijemos $H \in M(Q)$. Si σ es la acción de $\mathcal{O}(n)$ en $M(Q)$, luego componiendo a derecha por la acción a izquierda L_U con $U \in \mathcal{O}(n)$, obtenemos que $\tilde{\sigma}_U := \sigma \circ L_U$ manda Id en $U^T Q U$. De esta manera, tomando U tal que $U^T Q U = H$ tenemos que $T_H M(Q)$ coincide con la imagen de $D\tilde{\sigma}_U(\text{Id}) : T_{\text{Id}} \mathcal{O}(n) \rightarrow T_H M(Q)$. Como $\tilde{\sigma}_U(V) = V^T H V$ se tiene

$$D\tilde{\sigma}_U(\text{Id})\dot{U} = \dot{U}^T H + H\dot{U} = -\dot{U}H + H\dot{U} = [H, \dot{U}],$$

donde $\dot{U}^T = -\dot{U} \in T_{\text{Id}} \mathcal{O}(n)$. □

Teorema 3.4 (Flujo corchete doble). Sean N, Q simétricas.

1. La ecuación diferencial

$$\dot{H} = [H, [H, N]], \quad H(0)^T = H(0) = H_0 \quad (3.1)$$

define un flujo isoespectral en $\text{Sym}(n)$.

2. Si $f_N : M(Q) \rightarrow \mathbb{R}$ está dada por

$$f_N(H) = -\frac{1}{2} \|N - H\|^2,$$

entonces $H(t)$ solución de 3.1 satisface

$$\dot{H} = \text{grad } f_N(H),$$

donde el $\text{grad } f_N$ es respecto a la métrica Riemanniana de $M(Q)$ inducida como espacio homogéneo.

3. El flujo es completo y $H(t)$ converge a componentes conexas de puntos críticos. El conjunto de puntos críticos H_∞ , satisfacen $[H, N] = 0$.

4. Si $N = \text{diag}(\mu_1, \dots, \mu_n)$ con $\mu_i > \mu_j$ si $i < j$, entonces $H(t)$ converge, para $t \rightarrow \infty$, a matrices diagonales $\text{diag}(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)})$ siendo π una permutación y $\lambda_1, \dots, \lambda_n$ valores propios de H_0 .

Demostración. Dado que $[H, N]$ es antisimétrica cuando $H, N \in \text{Sym}(n)$, se tiene de la Proposición 2.60 que el flujo es isoespectral. Como $M(Q)$ es compacta el flujo es completo. Además, si $H(t)$ es una solución, se tiene

$$f_N(H) = -\frac{1}{2} \|N - H\|^2 = -\frac{1}{2} (\|N\|^2 + \|H\|^2) + \text{tr}(NH),$$

y por lo tanto

$$\frac{d}{dt} f_N(H(t)) = \text{tr}(N\dot{H}(t)) = \text{tr}(N[H, [H, N]]) = \text{tr}([N, H][N, H]) = \|[N, H]\|^2,$$

donde usamos la identidad $\text{tr}(A[B, C]) = \text{tr}([A, B]C)$. Se concluye que $\frac{d}{dt} f_N(H(t)) = \|[H, N]\|^2$ y por lo tanto $f_N(H(t))$ es estrictamente creciente. Como f_N es acotada resulta que es necesario que $\frac{d}{dt} f_N(H(t)) \rightarrow 0$. Es decir, $H(t)$ tiende a soluciones $H_\infty \in \text{Sym}(n)$ de $[H_\infty, N] = 0$.

Cuando $N = \text{diag}(\mu_1, \dots, \mu_n)$ con $\mu_i > \mu_j$ si $i < j$, $[H_\infty, N] = 0$ si y sólo si H_∞ es diagonal. Además como el flujo es isoespectral resulta que H_∞ es de la forma $\text{diag}(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)})$ con π una permutación y $\lambda_1, \dots, \lambda_n$ los valores propios de H_0 . Para probar la parte (b), observar que el gradiente de f_N en H es el único vector tangente en $T_H M(Q)$ que satisface

$$\langle \text{grad } f_N(H), [H, \Omega] \rangle = Df_N(H)\Omega, \quad \forall \Omega \in T_{\text{Id}}\mathcal{O}(n).$$

La métrica Riemanniana en $M(Q)$ está definida por

$$\langle [H, \Omega_1], [H, \Omega_2] \rangle_H := \text{tr}((\Omega_1|_{K^\perp})^T (\Omega_2|_{K^\perp}))$$

siendo $\Omega_i|_{K^\perp}$ la proyección en el espacio horizontal asociado. (Ver Apéndice C.6.)

Afirmación: $[H, N] \in T_{\text{Id}}\mathcal{O}(n)$, pertenece al espacio horizontal $K^\perp = \text{Ker } \pi(\text{Id})^\perp$, donde $\pi(U) = U^T H U$.

$\Omega \in T_{\text{Id}}\mathcal{O}(n)$ pertenece al núcleo del diferencial si y sólo si $[H, \Omega] = 0$, es decir, si conmuta con H . Luego, como $\text{tr}([H, N]^T \Omega) = \text{tr}(N[H, \Omega])$, resulta que $[H, N]$ es ortogonal a $K = \text{Ker } Df_N(\text{Id})$.

Ahora

$$\langle \text{grad } f_N, [H, \Omega] \rangle = \text{tr}(N[H, \Omega]) = \text{tr}([H, N]^T \Omega) = \text{tr}((\text{tr}([H, N]|_{K^\perp})^T \Omega|_{K^\perp}),$$

de donde, usando $Df_N(H)\dot{H} = \text{tr}(N^T \dot{H})$ se concluye que $\text{grad } f_N(H) = [H, [H, N]]$. \square

Comentario 3.5. Se puede probar que la función $f_N : M(Q) \rightarrow \mathbb{R}$ es siempre una función de Morse-Bott y por lo tanto las soluciones de

$$\dot{H} = [H, [H, N]]$$

convergen a un único punto crítico. Veamos en el caso en que N es diagonal y $Q = \text{diag}(\lambda_1, \dots, \lambda_n)$ con $\lambda_1 > \dots > \lambda_n$.

Lema 3.6. Sean $N = \text{diag}(\mu_1, \dots, \mu_n)$, $Q = \text{diag}(\lambda_1, \dots, \lambda_n)$ con $\lambda_1 > \dots > \lambda_n$ y $f_N : M(Q) \rightarrow \mathbb{R}$ dada por $f_N(H) = -\frac{1}{2}\|N - H\|^2$. Luego el hessiano de f_N en puntos críticos es no singular. La ecuación diferencial linealizada en un punto crítico H_∞ es

$$\dot{\zeta}_{ij} = -(\lambda_{\pi(i)} - \lambda_{\pi(j)})(\mu_i - \mu_j)\zeta_{ij} \quad i > j,$$

donde $H_\infty = \text{diag}(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)})$ y π una permutación.

Demostración. La linealización del campo $X(H) = [H, [H, N]]$ en $M(Q)$ en un punto crítico H_∞ es

$$DX(H_\infty)\zeta = H_\infty[\zeta, N] - [\zeta, N]H_\infty, \quad \zeta \in T_{H_\infty}M(Q),$$

luego, la ecuación diferencial linealizada resulta $\dot{\zeta} = H_\infty[\zeta, N] - [\zeta, N]H_\infty$. Como $\zeta \in T_{H_\infty}M(Q)$ y $H_\infty = \text{diag}(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)})$ se tiene $\zeta = [H_\infty, \Omega]$ con $\Omega^T = \Omega$, y por lo tanto $\zeta_{ij} = (\lambda_{\pi(i)} - \lambda_{\pi(j)})\omega_{ij}$ siendo $\Omega = (\omega_{ij})$. Observar que ζ es simétrica. Por lo tanto, derivando tenemos

$$\begin{aligned} \dot{\zeta}_{ij} &= \lambda_{\pi(j)}[\zeta, N]_{ij} - \lambda_{\pi(i)}[N, \zeta]_{ij} = \lambda_{\pi(j)}(\mu_i - \mu_j)\zeta_{ij} - \lambda_{\pi(i)}(\mu_j - \mu_i)\zeta_{ij} \\ &= -(\lambda_{\pi(i)} - \lambda_{\pi(j)})(\mu_i - \mu_j)\zeta_{ij}. \end{aligned}$$

Por lo tanto, la ecuación linealizada no tiene valores propios iguales a cero, y entonces el Hessiano es no singular. En particular, $\pi = \text{Id}$ es la única permutación con valores propios negativos y por lo tanto $H_\infty = Q$ es el atractor global. \square

Comentario 3.7. En el caso Q y N diagonales con valores propios distintos, concluimos que un abierto denso de $M(Q)$ converge a Q .

Comentario 3.8. Estos resultados nos permiten dar una prueba de la siguiente desigualdad:

Wielandt-Hoffman: $A, B \in \text{Sym}(n)$, entonces

$$\|A - B\|_F^2 \geq \sum_{i=1}^n (\lambda_i(A) - \lambda_i(B))^2,$$

siendo $\lambda_j(A), \lambda_j(B)$ los valores propios respectivos en orden decreciente.

En el caso de valores propios repetidos, la prueba de la desigualdad resulta de que los valores propios son funciones continuas de los coeficientes de la matriz.

Comentario 3.9. Cuando Q es de la forma $\begin{pmatrix} \text{Id}_k & 0 \\ 0 & 0 \end{pmatrix}$ el espacio homogéneo $M(Q)$ coincide con $\text{Grass}_{\mathbb{R}}(k, n)$ y entonces la función f_N es una función de Morse-Bott. Análogo para $St(k, n)$.

Comentario 3.10. La matriz N es un parámetro a elegir. Si nuestro objetivo es diagonalizar $A \in M(Q)$ basta tomar N matriz diagonal. Luego $H(t)$ tenderá a la matriz diagonal. El orden de los valores propios también está gobernado por N .

Comentario 3.11. Observar que la métrica Riemanniana homogénea no coincide con la métrica inducida.

\diamond **3.12.** Hallar expresión de $\text{grad } f_N$ para la métrica inducida.

3.1. Flujo en $\mathcal{O}(n)$

El método anterior nos sirvió para encontrar valores propios de una matriz simétrica. En esta sección queremos encontrar los vectores propios asociados. La función razonable a tomar es el levantado de f_N a $\mathcal{O}(n)$, o de la función $\text{tr}(NU^T H_0 H)$ con $H = U^T H_0 U$.

Sea $\varphi : \mathcal{O}(n) \rightarrow \mathbb{R}$ dada por $\varphi(U) = \text{tr}(NU^T H_0 U)$. Observar que

$$\|N - U^T H_0 U\|^2 = \|N\|^2 + \|H_0\|^2 - 2\varphi(U),$$

luego maximizar φ en $\mathcal{O}(n)$ es equivalente a minimizar $\|N - H\|$ con $H \in M(H_0)$.

Teorema 3.13. Sean $N, H_0 \in \text{Sym}(n)$

1. La ecuación diferencial $\dot{U} = H_0 U N - U N U^T H_0 U$ con $U(0) \in \mathcal{O}(n)$ es el flujo gradiente asociado a φ en $\mathcal{O}(n)$
2. El flujo es completo y converge a componentes de puntos de equilibrio $U_\infty \in \mathcal{O}(n)$, dados por los $U_\infty \in \mathcal{O}(n)$ tales que $[H, U_\infty^T H_0 U_\infty] = 0$.
3. En el caso H_0 con valores propios $\lambda_1 > \dots > \lambda_n$ y $N = \text{diag}(\mu_1, \dots, \mu_n)$ distintos, resulta $U(t) \rightarrow U_\infty$ que satisface

$$H_0 = U_\infty \text{diag}(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)}) U_\infty^T$$

con π permutación. En particular las columnas de U_∞ tienen los vectores propios de H_0 .

4. N y H_0 como en la parte anterior. El Hessiano de φ es no singular en puntos críticos. Hay un abierto denso de condiciones iniciales tales que $U(t) \rightarrow U_\infty$ satisfaciendo

$$H_0 = U_\infty \text{diag}(\lambda_1, \dots, \lambda_n) U_\infty^T.$$

Demostración. Derivando obtenemos $D\varphi(U) : T_U \mathcal{O}(n) \rightarrow \mathbb{R}$,

$$D\varphi(U)\dot{U} = \text{tr}(N\dot{U}^T H_0 U) + \text{tr}(NU^T H_0 \dot{U}).$$

Escribiendo $\dot{U} = U\Omega \in T_U\mathcal{O}(n)$ con Ω antisimétrica, se tiene

$$\begin{aligned} D\varphi(U)(U\Omega) &= \text{tr}(NU^T H_0 U \Omega) - \text{tr}(N\Omega U^T H_0 U) \\ &= \text{tr}(NU^T H_0 U \Omega) - \text{tr}(U^T H_0 U N \Omega) \\ &= \text{tr}([N, U^T H_0 U] \Omega). \end{aligned}$$

Por otro lado $D\varphi(U)(U\Omega) = \text{tr}(\text{grad } \varphi(U)^T U \Omega)$, siendo $\text{grad } \varphi(U) \in T_U\mathcal{O}(n)$ el gradiente de φ en $U \in \mathcal{O}(n)$. Luego resulta, del hecho que $U^T \text{grad } \varphi(U)$ es antisimétrica, que

$$U^T \text{grad } \varphi(U) = -[N, U^T H_0 U].$$

La completitud del flujo es por ser $\mathcal{O}(n)$ compacto, y el tipo de convergencia resulta del Teorema B.52 en el Apéndice, y comentarios posteriores. U_∞ es punto crítico si y sólo si $[N, U_\infty^T H_0 U_\infty] = 0$.

Para probar 3), observar que $U_\infty^T H_0 U_\infty$ es una matriz diagonal Λ . Por lo tanto, existe una cantidad finita de puntos críticos ($n!2^n$, donde 2^n viene de la elección v_i ó $-v_i$, siendo v_i vector propio). Entonces $H_0 = U_\infty \Lambda U_\infty^T$ es como el enunciado.

Para la última parte, linealizamos la ecuación diferencial en el entorno de un punto crítico $X(U) = -U[N, U^T H_0 U]$, con $U \in \mathcal{O}(n)$ luego

$$DX(U)\zeta = -U[N, \zeta^T H_0 + U^T H_0 \zeta] \in T_U\mathcal{O}(n),$$

para $\zeta = U\Omega \in T_U\mathcal{O}(n)$ con Ω antisimétrica. Luego la ecuación diferencial linealizada en función de Ω es

$$\dot{\Omega} = -[N, [U^T H_0 U, \Omega]] \in T_{\text{Id}}\mathcal{O}(n).$$

Tomando coeficientes y observando que N y $U_\infty^T H_0 U_\infty = \Lambda$ son diagonales, resulta

$$\dot{\omega}_{ij} = -(\lambda_{\pi(i)} - \lambda_{\pi(j)})(\mu_i - \mu_j)\omega_{ij} \quad i > j,$$

siendo $\Omega = (\omega_{ij}) \in T_{\text{Id}}\mathcal{O}(n)$.

Argumentando al igual que en el resultado anterior concluimos el resultado. \square

◇ **3.14.** Probar que el gradiente de la función $g : M(Q) \rightarrow \mathbb{R}$ dada por

$$g(H) = \|H - \text{diag } H\|^2$$

resulta

$$\text{grad } g(H) = [H, [H, \text{diag } H]].$$

Los puntos de equilibrio asociados son H_∞ tales que $[H_\infty, \text{diag } H_\infty] = 0$.

¿Cuál es la ecuación diferencial asociada al levantado?

3.2. Flujos Toda

En muchas ocasiones, para encontrar valores propios de matrices es conveniente transformar nuestra matriz a una con una estructura que nos permita ser más eficientes. Un ejemplo de esto, en el caso simétrico, es transformar nuestra matriz a ser tridiagonal (forma de Hessenberg). Esta transformación se puede hacer a "bajo" costo.

Definición 3.15 (Matriz de Hessenberg). Una matriz $A = (a_{ij})$ simétrica es de Hessenberg, o de Jacobi, si $a_{ij} = 0$ siempre que $|i - j| \geq 2$.

Una propiedad agradable de los flujos de corchete doble es la siguiente.

Lema 3.16. Sean $N = \text{diag}(1, \dots, n)$ y H de Jacobi, entonces $[H, [H, N]]$ es de Jacobi. En particular el flujo $H(t)$ asociado a $\dot{H} = [H, [H, N]]$ con $H(0)$ de Jacobi es un flujo isoespectral en las matrices de Jacobi.

Demostración. Observar que en general, si $H = (h_{ij})$ resulta que $[H, N]_{ij} = (j - i)h_{ij}$. Por lo tanto

$$A_{ij} = [H, [H, N]]_{ij} = \sum_{k=1}^n h_{ik}(j-k)h_{kj} - \sum_{k=1}^n (k-i)h_{ik}h_{kj} = \sum_{k=1}^n (i+j-2k)h_{ik}h_{kj}. \quad (3.2)$$

Como H es de Jacobi, entonces se tiene que $h_{ij} = 0$ si $|i - j| \geq 2$, por lo que en principio $A_{ij} = 0$ si $|i - j| > 2$, dado que si $|i - j| \geq 3$ entonces $\max_{k \in \{1, \dots, n\}} |i - k|, |j - k| \geq 2$.

Además, si $j = i + 2$, entonces la única posible solución no cero sería $k = \frac{i+j}{2}$. Luego (3.2) tomaría el valor 0. Luego concluimos que $[H, [H, N]]$ es de Jacobi. \square

Comentario 3.17. Observar que $[H, N] = \begin{pmatrix} 0 & h_{12} & \dots & h_{1n} \\ -h_{21} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & h_{1(n-1)} \\ -h_{n1} & \dots & -h_{n(n-1)} & 0 \end{pmatrix}$, luego,

si H es de Jacobi, $H = \begin{pmatrix} a_1 & b_1 & 0 & \dots & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots \\ 0 & b_2 & a_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{n-1} \\ 0 & \dots & 0 & b_{n-1} & a_n \end{pmatrix}$, entonces

$$[H, N] = \begin{pmatrix} 0 & b_1 & 0 & \dots & 0 \\ -b_1 & 0 & b_2 & \ddots & \vdots \\ 0 & -b_2 & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{n-1} \\ 0 & \dots & 0 & -b_{n-1} & 0 \end{pmatrix}.$$

En particular, $[H, N] = H_u - H_l$ siendo H_u y H_l las matrices triangulares superior e inferior asociadas a H .

Definición 3.18 (Flujo Toda). La ecuación diferencial $\dot{H} = [H, H_u - H_l]$ se le denomina Flujo Toda.

Esta ecuación tiene la siguiente interpretación física definida por Toda en la década del 60.

Consideremos n partículas x_1, \dots, x_n en una recta y sean y_1, \dots, y_n las respectivas velocidades. Supongamos que el sistema de partículas está gobernado por el Hamiltoniano

$$H(x, y) = K(y) + V(x) = \frac{1}{2} \sum_{k=1}^n y_k^2 + \sum_{i=1}^n \exp(x_k - x_{k+1}),$$

a $K(y)$ se lo denomina la energía cinética del sistema, mientras que a $V(x)$ se lo denomina la energía potencial (donde agregamos las partículas ficticias $x_0 = -\infty$ y $x_{n+1} = +\infty$, ambos con energía potencial cero).

Desde el punto de vista matemático, permitimos que no hayan colisiones, asumiendo de esta manera que las partículas pueden cruzarse.

Fijamos una condición inicial $(x(0), y(0)) \in \mathbb{R}^{2n}$ y sea $(x(t), y(t)) \in \mathbb{R}^{2n}$ el estado del sistema en tiempo $t \in \mathbb{R}$. Luego el sistema está determinado por las ecuaciones de Hamilton-Jacobi

$$\begin{cases} \dot{x}_k = \frac{\partial H}{\partial y_k} = y_k \\ \dot{y}_k = -\frac{\partial H}{\partial x_k} = \exp(x_{k-1} - x_k) - \exp(x_k - x_{k+1}) \end{cases} \quad (3.3)$$

Estas ecuaciones implican que la energía $H(x(t), y(t))$ es una cantidad que se conserva.

El problema fundamental de la teoría de la mecánica Hamiltoniana es encontrar una *integral del sistema*, i.e. encontrar el mapa $t \mapsto (x(t), y(t))$. Observar que el sistema es invariante por traslaciones, es decir

$$H(x + \vec{c}, y) = H(x, y), \quad \vec{c} = (c, \dots, c) \in \mathbb{R}^n.$$

Esto implica que el centro de masa se mueve de manera uniforme.

Flasche (74') utilizo esta observación para hacer un cambio de coordenadas, trasladar al centro de masa (trabajar con las diferencias $x_k - x_{k+1}$), eliminar exponenciales y ajustar constantes, $a_k = \frac{y_k}{2}$, $b_k = \frac{1}{2} \exp\left(\frac{x_k - x_{k+1}}{2}\right)$. De donde resulta,

$$\dot{a}_k = \frac{1}{2} \dot{y}_k = \frac{1}{2} (\exp(x_{k-1} - x_k) - \exp(x_k - x_{k+1})) = \frac{1}{2} ((2b_{k-1})^2 - (2b_k)^2) = 2(b_{k-1}^2 - b_k^2),$$

$$\dot{b}_k = \frac{1}{2} \exp\left(\frac{x_k - x_{k+1}}{2}\right) \frac{\dot{x}_k - \dot{x}_{k+1}}{2} = b_k \left(\frac{y_k - y_{k+1}}{2}\right) = b_k(a_k - a_{k+1}).$$

Es decir que en estas nuevas coordenadas el sistema 3.3 es equivalente a

$$\begin{cases} \dot{a}_k = 2(b_{k-1}^2 - b_k^2) \\ \dot{b}_k = b_k(a_k - a_{k+1}) \end{cases} \quad (3.4)$$

Proposición 3.19. Sean $H(0) = \begin{pmatrix} a_1 & b_1 & 0 & \dots & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots \\ 0 & b_2 & a_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & b_{n-1} \\ 0 & \dots & 0 & b_{n-1} & a_n \end{pmatrix}$ y $N = \text{diag}(1, \dots, n)$.

La ecuación diferencial de Toda $\dot{H} = [H, [H, N]]$ está dada por la ecuación diferencial (3.4).

Demostración. Recordar del Lema que podemos suponer

$$\dot{H} = \begin{pmatrix} \dot{a}_1 & \dot{b}_1 & 0 & \dots & 0 \\ \dot{b}_1 & \dot{a}_2 & \dot{b}_2 & \ddots & \vdots \\ 0 & \dot{b}_2 & \dot{a}_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \dot{b}_{n-1} \\ 0 & \dots & 0 & \dot{b}_{n-1} & \dot{a}_n \end{pmatrix}.$$

Luego de la ecuación 3.2 resulta que

$$\dot{a}_i = \sum_{k=i-1, i+1} (2i-2k)h_{ik}h_{ki} = 2(b_{i-1}^2 - b_i^2).$$

Por otro lado,

$$\dot{b}_i = \dot{H}_{i(i+1)} = \sum_{k=i, i+1} (2i+1-2k)h_{ik}h_{k(i+1)} = a_i b_i - b_i a_{i+1} = b_i(a_i - a_{i+1}).$$

□

Comentario 3.20. Sea $Jac(\lambda_1, \dots, \lambda_n)$ el conjunto de matrices $n \times n$ de Jacobi con valores propios $\lambda_1 \geq \dots \geq \lambda_n$. La topología y geometría de este conjunto no es fácil de describir.

Tomei (84') probó que si $\lambda_1 > \dots > \lambda_n$ entonces $Jac(\lambda_1, \dots, \lambda_n)$ es una variedad diferenciable de dimensión $n-1$, compacta.

Además si $n=3$, se tiene que es una superficie de Riemann de género 2.

Del comentario anterior y lo estudiado para la ecuación (3.1) se concluye el siguiente resultado.

Teorema 3.21 (Flujo Toda). 1. El flujo de Toda $\dot{H} = [H, H_u - H_l]$ es el flujo isoespectral en el conjunto de matrices de Jacobi.

2. La solución $H(t)$ existe para todo $t \in \mathbb{R}$ y converge a una diagonal cuando $t \rightarrow \pm\infty$.
3. Sea $N = \text{diag}(1, \dots, n)$ y $\lambda_1 > \dots > \lambda_n$. El flujo de Toda en la variedad isoespectral $Jac(\lambda_1, \dots, \lambda_n)$ es el flujo gradiente de la función

$$f_N : Jac(\lambda_1, \dots, \lambda_n) \rightarrow \mathbb{R},$$

dada por $f_N(H) = \frac{-1}{2} \|N - H\|^2$, para la métrica inducida como subvariedad del espacio homogéneo $M(H_0)$.

Comentario 3.22. Del corolario resulta que podemos dar una interpretación física al sistema Hamiltoniano estudiado. Como $b_k \rightarrow 0$ esto significa que $x_k - x_{k+1} \rightarrow -\infty$ y por lo tanto cada partícula se comporta de manera independiente y con velocidades asintóticas constantes, manteniendo el centro de masa con velocidad uniforme.

3.3. Flujos isoespectrales y ecuación de Riccati

Recordar que la variedad $M(Q) = \{UQU^T : U \in \mathcal{O}(n)\}$ para $Q = \text{diag}(\lambda_1 \text{Id}_{n_1}, \dots, \lambda_r \text{Id}_{n_r})$ con $\lambda_1 > \dots > \lambda_r$ es difeomorfo al espacio homogéneo $\mathcal{O}(n)/\mathcal{O}(n_1) \times \dots \times \mathcal{O}(n_r)$. La anterior identificación la hacíamos considerando la acción por conjugación de $\mathcal{O}(n)$ en $M(Q)$.

Lema 3.23. $M(Q)$ es difeomorfo al espacio bandera $\text{Flag}(n_1, \dots, n_r)$. La identificación está dada por

$$UQU^T \mapsto (\langle U_1^T \rangle, \langle U_1^T U_2^T \rangle, \dots, \langle U_1^T \dots U_r^T \rangle)$$

siendo U_1, \dots, U_r tales que $U_i \in \mathbb{R}^{n_i \times n}$ y $U = \begin{pmatrix} U_1 \\ \vdots \\ U_r \end{pmatrix}$.

◇ **3.24.** Probar el lema anterior.

Con esta identificación podemos preguntarnos cuál es el flujo asociado en $\text{Flag}(n_1, \dots, n_r)$ al flujo isoespectral $\dot{H} = [H, [H, N]]$. Daremos una respuesta para el caso $r = 2$, donde $\text{Flag}(k, n-k)$ coincide con $\text{Grass}(k, n)$.

Toda matriz $N \in \mathbb{R}^{n \times n}$, induce un flujo en $\text{Grass}(k, n)$, dado por

$$\Phi_N(t, V) = e^{tN}V,$$

siendo $e^{tN}V$ la acción natural de $\mathbb{G}l(n, \mathbb{R})$ en $\text{Grass}(k, n)$. Diremos que este flujo en $\text{Grass}(k, n)$ es *inducido linealmente* por N .

Teorema 3.25. Sean $N \in \text{Sym}(n)$, $Q = \text{diag}(\lambda_1 \text{Id}_k, \lambda_2 \text{Id}_{n-k})$ con $\lambda_1 > \lambda_2$ y $1 \leq k < n-1$. El flujo generado por $\dot{H} = [H, [H, N]]$ inducido a través de la identificación del Lema 3.23, es un flujo isoespectral en $\text{Grass}(k, n)$ inducido linealmente por $(\lambda_1 - \lambda_2)N$.

Demostración. Sea $H(t) = U(t)QU(t)^T$ solución de $\dot{H} = [H, [H, N]]$ con condición inicial $H_0 = U_0QU_0^T$. Nuestro objetivo es probar que $\varphi(H(t)) = e^{t(\lambda_1 - \lambda_2)N}V_0 \in \text{Grass}(k, n)$ siendo $V_0 = \varphi(H_0)$ y $\varphi : M(Q) \rightarrow \text{Grass}(k, n)$ el difeomorfismo $\varphi(UQU^T) = \langle U_1^T \rangle$.

Del Teorema 3.13 probamos que $U(t) \in \mathcal{O}(n)$ satisface la condición

$$\dot{U} = U(U^T QUN - NU^T QU).$$

Sea $X(t) = U(t)^T$, de donde tenemos $\varphi(H(t)) = \langle X(t)_1^T \rangle$. Con este cambio, tenemos que $X(t)$ satisface la siguiente ecuación

$$\dot{X} = NXQ - XQX^T NX. \quad (3.5)$$

Consideremos ahora $S(t) \in \mathbb{R}^{n \times n}$, con $S(0) = \text{Id}$ que sea solución de

$$\dot{S} = X^T NX [(\lambda_1 - \lambda_2)S - QS] + QX^T NX S. \quad (3.6)$$

Escribiendo $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$ por bloques, si suponemos que $S_{21}(0) = 0$ entonces se tiene que $\dot{S}_{21} = 0$ y por lo tanto la solución es de la forma

$$S = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix}.$$

Observar que entonces $X(t)S(t) = e^{t(\lambda_1 - \lambda_2)N} X(0)$. En efecto, sea $Y(t) = X(t)S(t)$, entonces $\dot{Y} = \dot{X}S + X\dot{S}$. Por lo tanto

$$\begin{aligned} \dot{Y} &= NXQS - XQX^T NX S + XX^T NX [(\lambda_1 - \lambda_2)S - QS] + XQX^T NX S \\ &= (\lambda_1 - \lambda_2)NXS = (\lambda_1 - \lambda_2)NY. \end{aligned}$$

Luego, $Y(t)$ satisface $\dot{Y} = (\lambda_1 - \lambda_2)NY$ con condición inicial $Y(0) = X(0)S(0) = X(0)$. Entonces, por unicidad de soluciones resulta

$$Y(t) = e^{t(\lambda_1 - \lambda_2)N} X(0).$$

Recordar que $\varphi(H(t)) = \langle X(t)_1^T \rangle$. De lo anterior resulta que $S(t) \in \mathbb{G}_1(n, \mathbb{R})$ y por lo tanto $\langle X(t)_1^T \rangle = \langle S_{11}(t)^T X(t)_1^T \rangle$.

Por lo tanto, el espacio generado por las primeras k columnas de $X(t)S(t)$ coincide con

$$e^{t(\lambda_1 - \lambda_2)N} X_1(0) = e^{t(\lambda_1 - \lambda_2)N} V_0.$$

□

Comentario 3.26. Si recordamos la ecuación diferencial de Riccati

$$\dot{K} = A_{21} + A_{22}K - KA_{11} - KA_{12}K,$$

$K \in \mathbb{R}^{(n-k) \times k}$, donde $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, se tiene que $e^{tA}V_0 = \left[\begin{pmatrix} Id_k \\ K(t) \end{pmatrix} \right] \in \text{Grass}(k, n)$.

Luego, $H(t)$ induce la ecuación de Riccati $K(t)$ con $A = (\lambda_1 - \lambda_2)N$

Otras aplicaciones

En este capítulo veremos otras aplicaciones de problemas de optimización en variedades.

4.1. Programación lineal en el caso convexo

El problema de Programación Lineal a resolver consiste en lo siguiente.

PROBLEMA (Programación lineal): *Hallar $x^* \in \mathbb{R}^n$ tal que:*

$$x^* \in \arg \min_{x \in F} c^T x,$$

donde $c \in \mathbb{R}^n$ no nulo, y el conjunto factible F es la combinación convexa de m vectores $v_i \in \mathbb{R}^n$ (vértices de F):

$$F = \text{Conv}(v_1, v_2, \dots, v_m) := \left\{ v = \sum_{i=1}^m \alpha_i v_i, \alpha_i \geq 0, \sum_{i=1}^m \alpha_i = 1 \right\} \subset \mathbb{R}^n$$

Este conjunto factible F es convexo y compacto. En lo que sigue asumimos que $c^T v_i \neq c^T v_j, \forall i \neq j$; con lo cual se puede afirmar que el problema tiene solución única y además esta se alcanza en uno de los vértices, al cual denotamos v_s .

Lo anterior implica en particular que, si se conocen los valores de la función objetivo en todos los vértices $c^T v_i$, es fácil saber en que vértice se alcanza el óptimo, previa comparación de los valores. Por lo tanto, en el caso de interés, donde no resulta eficiente hallar todos los valores $c^T v_i$, se tendrán mas vértices que la dimensión del espacio: $m > n$.

4.1.1. Ecuación de Brockett

Sean las matrices diagonal $N = \text{diag}(c^T v_1, \dots, c^T v_m) \in \mathbb{R}^{m \times m}$ y $Q = \text{diag}(1, \dots, 0) \in \mathbb{R}^{m \times m}$. Consideremos la siguiente ecuación diferencial en el espacio de las matrices simétricas $\text{Sym}(m)$:

$$\begin{aligned} \dot{H} &= [H, [H, N]] = H^2 N + N H^2 - 2 H N H \\ H(0) &= \Theta^T Q \Theta \end{aligned} \quad (4.1)$$

Por el Teorema 3.13, se puede afirmar que: para casi toda matriz $\Theta \in O(m)$ ortogonal, la solución $H(t)$ del flujo anterior converge a $H_\infty = \text{diag}(0, \dots, 1, \dots, 0) \in \mathbb{R}^{m \times m}$ cuando $t \rightarrow \infty$, donde el 1 se encuentra en la posición óptima s . Es decir que, conociendo H_∞ , se conoce el vértice v_s donde se alcanza el óptimo.

Desventajas del método

1. El flujo evoluciona en el espacio de las matrices simétricas $\text{Sym}(m)$, el cual tiene dimensión $\frac{m(m+1)}{2} \gg n$; mientras que la región factible F del problema original está contenida en un espacio de dimensión n .
2. El resultado anterior no proporciona una caracterización de los $\Theta \in O(m)$ iniciales para los cuales se garantiza convergencia.
3. Para definir la matriz N hay que conocer los valores de la función objetivo en todos los vértices: $c^T v_i$. Como el óptimo se alcanza en uno de estos vértices, conocer N equivale entonces a conocer la solución del problema (previa comparación de los valores).

En lo que sigue vamos a ver cómo superar las primeras dos desventajas.

4.1.2. Ecuación de Brockett simplificada (en la esfera)

La idea es plantear un problema equivalente al original, consistente en maximizar una función de Rayleigh en la esfera $S^{m-1} \subset \mathbb{R}^m$, de dimensión $m - 1$. En particular esto permite reducir la dimensión del espacio de trabajo, superando la primera desventaja de la ecuación de Brockett dada en (4.1).

Definición de espacios y funciones

Definición 4.1 (Simplex). El *simplex* de dimensión $m - 1$ es:

$$\Delta_{m-1} := \left\{ y \in \mathbb{R}^m, y_i \geq 0, \sum_{i=1}^m y_i = 1 \right\} \subset \mathbb{R}^m$$

Se definen la función

$$f: S^{m-1} \rightarrow \Delta_{m-1}, \quad f(\theta) := (\theta_1^2, \dots, \theta_m^2) \in \mathbb{R}^m$$

y la función lineal

$$T: \Delta_{m-1} \rightarrow \text{Conv}(v_1, v_2, \dots, v_m), \quad T(y) = \sum_{i=1}^m y_i v_i.$$

Observar que la matriz asociada a T es la matriz $[v_1, \dots, v_m] \in \mathbb{R}^{n \times m}$. Definimos finalmente $\pi_T = T \circ f: S^{m-1} \rightarrow \text{Conv}(v_1, v_2, \dots, v_m)$ y la función objetivo a optimizar sobre la esfera: $\lambda \circ \pi_T: S^{m-1} \rightarrow \mathbb{R}$, donde $\lambda(x) = c^T x$ es la función lineal inicial.

La Figura 4.1 muestra un diagrama con los conjuntos y funciones definidas anteriormente.

Función objetivo

Lema 4.2. En la esfera S^{m-1} , la función $\lambda \circ \pi_T$ se puede expresar como el siguiente cociente de Rayleigh: $(\lambda \circ \pi_T)(\theta) = \frac{\theta^T N \theta}{\theta^T \theta}$.

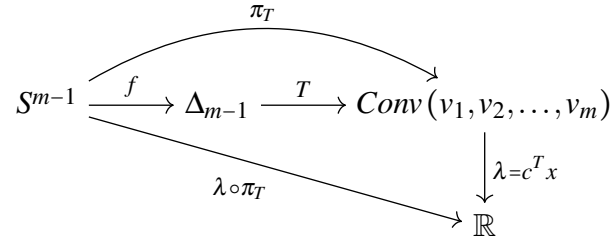


Figura 4.1: Diagrama de espacios y funciones

Demostración.

$$\begin{aligned}
 (\lambda \circ \pi_T)(\theta) &= \lambda(\pi_T(\theta)) = c^T \pi_T(\theta) = c^T T(\theta_1^2, \dots, \theta_m^2) = \dots \\
 &= c^T \sum_{k=1}^m \theta_k^2 \vec{v}_k = \sum_{i=1}^n c_i \left(\sum_{k=1}^m \theta_k^2 \vec{v}_k \right)_i = \sum_{i=1}^n c_i \left(\sum_{k=1}^m \theta_k^2 v_{ki} \right) = \sum_{k=1}^m \theta_k \left(\sum_{i=1}^n c_i v_{ki} \right) \theta_k = \dots \\
 &= \sum_{k=1}^m \theta_k (c^T \vec{v}_k) \theta_k = \theta^T N \theta = \frac{\theta^T N \theta}{\theta^T \theta}
 \end{aligned}$$

□

Definimos en la esfera S^{m-1} una métrica Riemanniana dada por el producto interno:

$$\langle \theta, y \rangle = 2\theta^T y, \quad \theta, y \in T_x S^{m-1}$$

Flujo de Brockett en la esfera

Teorema 4.3 (Flujo de Brockett en S^{m-1}). *Sea $N = \text{diag}(c^T v_1, \dots, c^T v_m) \in \mathbb{R}^{m \times m}$ como antes. Entonces:*

1. *El flujo de gradiente de $\lambda \circ \pi_T$ en la esfera S^{m-1} , y respecto a la métrica anterior, es:*

$$\begin{aligned}
 \dot{\theta} &= \text{grad}(\lambda \circ \pi_T) = (N - \theta^T N \theta \text{Id}) \theta \\
 \theta(0) &\in \mathbb{R}^m
 \end{aligned} \tag{4.2}$$

2. *El flujo anterior tiene exactamente 2^m puntos de equilibrio, dados por: $\pm e_1, \dots, \pm e_m$, donde e_i es el i -ésimo vector canónico.*

3. Sea $\dot{y} = A_{\pm i}y$ la linealización del flujo de gradiente en $\pm e_i$. Entonces los $m-1$ valores propios de $A_{\pm i}$ son (los valores no nulos):

$$c^T(v_1 - v_i), \dots, c^T(v_{i-1} - v_i), c^T(v_{i+1} - v_i), \dots, c^T(v_m - v_i)$$

Además: el índice óptimo s es el único tal que $e_{\pm s}$ es asintóticamente estable ($A_{\pm s}$ es definida negativa).

4. Sea $X \cong S^{m-2}$ la subvariedad diferenciable de S^{m-1} , de codimensión uno, definida como:

$$X = \{\theta \in S^{m-1}, \theta_s = 0\}$$

Si $\theta(0) \in S^{m-1} \setminus X$, la solución $\theta(t)$ del flujo de Brockett simplificado converge exponencialmente al atractor estable $\pm e_s$. Mas aún, $\pi_T(\theta(t))$ converge exponencialmente al óptimo v_s del problema de programación lineal.

□

El último punto de este resultado permite superar la segunda desventaja de la ecuación de Brockett dada en (4.1).

4.1.3. Ecuación de Brockett simplificada (para PL en el simplex)

Consideremos ahora el caso en que la región factible $F = \text{Conv}(v_1, v_2, \dots, v_m) \subset \mathbb{R}^n$ del problema de Programación Lineal es el conjunto simplex $\Delta_{m-1} \subset \mathbb{R}^m$. En particular se tiene $m = n$.

En este caso $T = \text{Id}$ o, equivalentemente, $v_i = e_i, \forall i$. Por lo tanto:

$$N = \text{diag}(c^T e_1, \dots, c^T e_n) = \text{diag}(c_1, \dots, c_n) \quad \text{y} \quad \theta^T N \theta = \sum_{k=1}^n c_k \theta_k^2$$

El flujo de Brockett (en S^{n-1}) es entonces:

$$\dot{\theta}_i = \left(c_i - \sum_{k=1}^n c_k \theta_k^2 \right) \theta_i, \forall i = 1, \dots, n \quad (4.3)$$

Con $\sum_{i=1}^n \theta_i^2 = 1$ por estar trabajando en la esfera.

Introduciendo el cambio de variable $x_i = \theta_i^2$, el flujo anterior se puede plantear sobre el simplex, en lugar de la esfera. Derivando se obtiene $\dot{x}_i = 2\theta_i \dot{\theta}_i$ y reemplazando se llega al flujo en x :

$$\dot{x}_i = 2 \left(c_i - \sum_{k=1}^n c_k x_k \right) x_i, \forall i = 1, \dots, n \quad (4.4)$$

Como $\sum_{i=1}^n \theta_i^2 = 1$ entonces $\sum_{i=1}^n x_i = 1$. Además $x_i = \theta_i^2 \geq 0$, por lo que se concluye que el flujo anterior se da en el simplex: $x \in \Delta_{n-1}$.

En este caso el conjunto $X \subset S^{n-1}$, de condiciones iniciales excepcionales que garantizan convergencia, es llevado por el cambio de variables a un subconjunto de $\partial\Delta^{n-1}$.

Corolario 4.4. *La ecuación (4.4) define un flujo en Δ_{n-1} tal que: si $x(0) \in \Delta_{n-1} \setminus \partial\Delta_{n-1} = \text{int}(\Delta_{n-1})$, entonces la solución $x(t)$ converge exponencialmente a e_s , solución del problema de Programación Lineal (en el simplex Δ_{n-1}).*

Observaciones

- El flujo en x es una ecuación del tipo Lotka-Volterra
- $x(t)$ converge a e_s desde el interior de Δ_{n-1} , por lo que es un método de tipo “punto interior”.
- Si en $\text{int}(\Delta_{n-1})$ se define la métrica dada por el producto interno: $\langle \theta, y \rangle = \sum_{i=1}^n \frac{\theta_i y_i}{x_i}$, entonces el anterior flujo en x es flujo de gradiente de $c^T x$ en $\text{int}(\Delta_{n-1})$.

4.2. Programación Lineal en un Polítopo

Dada $A \in \mathbb{R}^{m \times n}$ y $b \in \mathbb{R}^m$, se considera el conjunto tipo Polítopo (Polytope) dado por:

$$C = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\} \subset \mathbb{R}^n$$

Este conjunto es convexo y en lo que sigue se asume que además es compacto y con interior no vacío¹. Como se verá más adelante, el simplex Δ_{n-1} es un caso particular de este conjunto C .

Se considera ahora el problema de optimización, mas general que el de Programación Lineal, consistente en hallar x^* tal que: $x^* \in \max_{x \in C} \phi(x)$, para cierta función $\phi : C \rightarrow \mathbb{R}$ no necesariamente lineal.

En lo que sigue se define una métrica en $\overset{\circ}{C}$ y se obtiene una expresión para el flujo de gradiente de ϕ en $\overset{\circ}{C}$ respecto a dicha métrica: $\dot{x} = \text{grad}(\phi(x))$. Esto permite tener funciones $x(t)$ tales que $\phi(x(t)) \rightarrow \phi(x^*)$

4.2.1. Gradiente de ϕ en $\overset{\circ}{C}$

Dado $x \in \overset{\circ}{C}$, el espacio tangente a $\overset{\circ}{C}$ en dicho punto es:

$$T_x \overset{\circ}{C} = T_x \{x \in \mathbb{R}^n : Ax = b\} = \{\theta \in \mathbb{R}^n : A\theta = \vec{0}\} = \ker(A)$$

Por otro lado, la matriz diagonal $D(x) = \text{diag}(x_1, \dots, x_n)$ es definida positiva para todo $x \in \overset{\circ}{C}$, por lo que siempre es invertible. Esto permite definir un producto interno en $T_x \overset{\circ}{C}$ dado por:

$$\langle\langle \theta, y \rangle\rangle = \theta^T D(x)^{-1} y, \quad \forall \theta, y \in T_x \overset{\circ}{C}$$

El operador $\text{grad} \phi$, respecto a la métrica inducida por $\langle\langle \cdot, \cdot \rangle\rangle$, es tal que:

1. $\text{grad} \phi(x) \in T_x \overset{\circ}{C}, \forall x \in T_x \overset{\circ}{C}$ y
2. $\langle\langle \text{grad} \phi(x), \theta \rangle\rangle = \nabla \phi(x)^T \theta, \forall \theta \in T_x \overset{\circ}{C}$

Usando la definición del producto interno, la segunda condición es equivalente a:

$$(\text{grad} \phi(x)^T D(x)^{-1} - \nabla \phi(x)^T) \theta = 0$$

si y sólo si

$$(D(x)^{-1} \text{grad} \phi(x) - \nabla \phi(x))^T \theta = 0, \quad \forall \theta \in \text{Ker}(A)$$

¹Su interior esta dado por: $\overset{\circ}{C} = \{x \in \mathbb{R}^n : Ax = b, x > 0\}$

Es decir:

$$D(x)^{-1} \text{grad } \phi(x) - \nabla \phi(x) \in \text{Ker}(A)^T = \text{Im}(A^T)$$

Por lo tanto debe existir $\lambda \in \mathbb{R}^m$ tal que:

$$D(x)^{-1} \text{grad } \phi(x) - \nabla \phi(x) = A^T \lambda \Leftrightarrow A \text{grad } \phi(x) - AD(x) \nabla \phi(x) = AD(x) A^T \lambda$$

Si asumimos que el rango de A es máximo: $\text{rg}(A) = m < n$, este λ será único. Para hallarlo se utiliza la primer condición que debe cumplir $\text{grad } \phi$: $\text{grad } \phi(x) \in \text{Ker}(A)$. Reemplazando esto en la última expresión se obtiene:

$$\vec{0} - AD(x) \nabla \phi(x) = AD(x) A^T \lambda \Leftrightarrow \lambda = - (AD(x) A^T)^{-1} AD(x) \nabla \phi(x)$$

4.2.2. Flujo de gradiente de ϕ en \mathring{C}

El flujo de gradiente de ϕ en \mathring{C} , con respecto a la métrica anterior, es entonces:

$$\dot{x} = \text{grad } \phi(x) = \left[Id - D(x) A^T (AD(x) A^T)^{-1} A \right] D(x) \nabla \phi(x)$$

Este se conoce como flujo de Faybusovich.

4.2.3. Flujo en el Simplex

En el caso particular en que $A = (1, \dots, 1) \in \mathbb{R}^{1 \times n}$ y $b = 1$, el Politopo es el simplex: $C = \{x \in \mathbb{R}^n, \sum_{i=1}^n x_i = 1, x_i \geq 0\} = \Delta_{n-1}$. Además $AD(x) = x^T$ y $D(x)A^T = x$. Por lo tanto el flujo de ϕ en el Simplex es:

$$\begin{aligned} \dot{x} &= D(x) \nabla \phi(x) - x x^T \nabla \phi(x) \Leftrightarrow \\ \dot{x}_i &= \left[\frac{\partial \phi}{\partial x_i}(x) - \sum_{k=1}^n x_k \frac{\partial \phi}{\partial x_k}(x) \right] x_i, \forall i = 1, \dots, n \end{aligned}$$

Función objetivo lineal

Cuando ϕ es lineal: $\phi(x) = c^T x$, se obtiene el flujo en Δ_{n-1} dado por la ecuación de Brockett (4.4) (a menos de un factor constante 2):

$$\dot{x}_i = \left(c_i - \sum_{k=1}^n c_k x_k \right) x_i, \forall i = 1, \dots, n$$

Función objetivo cuadrática

Cuando ϕ es cuadrática: $\phi(x) = \frac{1}{2} \sum_{k=1}^n c_k x_k^2$, su gradiente libre tiene $(\nabla \phi(x))_i = c_i x_i$, por lo que el flujo en Δ_{n-1} es:

$$\dot{x}_i = \left(c_i x_i - \sum_{k=1}^n c_k x_k^2 \right) x_i, \forall i = 1, \dots, n$$

Este se conoce como flujo de Karmarkar.

4.3. Flujos en $\mathcal{O}(n)$ y Graph Matching

En esta sección daremos una descripción del problema denominado *Graph Matching*, y su relación con flujos en $\mathcal{O}(n)$.

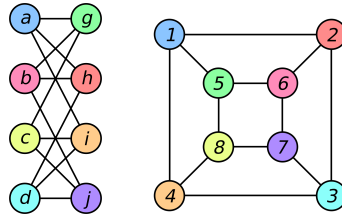
Definición 4.5 (Grafo). Un *grafo* $G = (V, E)$ consiste en un conjunto de vértices $V = \{v_1, v_2, \dots, v_n\}$ y un conjunto de aristas E , que son pares no ordenados de V .

Intuitivamente, un grafo es un conjunto de puntos unidos por aristas, por lo que hay muchas formas de “dibujarlo”. En particular, cualquier re-ordenación de los nodos, da lugar al mismo dibujo. El concepto de isomorfismo captura este concepto:

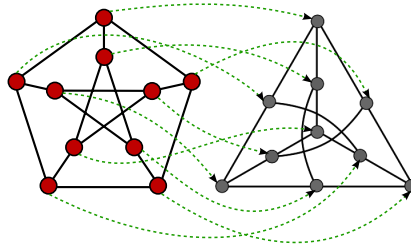
Definición 4.6 (Grafos isomorfos). Decimos que dos grafos $G_1 = (V_1, E_1)$ y $G_2 = (V_2, E_2)$ son isomorfos si y sólo si existe una biyección $\varphi : V_1 \rightarrow V_2$ que preserve la estructura. Esto es, que $(v_i, v_j) \in E_1 \iff (\varphi(v_i), \varphi(v_j)) \in E_2$

Definición 4.7 (Automorfismo). Un isomorfismo de un grafo en sí mismo se denomina *automorfismo*, y el conjunto de automorfismos de un grafo forma un grupo con la composición.

Ejemplos 4.8. Los grafos de la siguiente figura son isomorfos, siendo φ la que asigna a un vértice de un grafo, el vértice del mismo color en el otro grafo.



Ejemplos 4.9. Los grafos de la siguiente figura son isomorfos, con la biyección representada con las flechas punteadas.



PROBLEMA (Isomorfismo de grafos): *Dados dos grafos, decidir si son isomorfos.*

El problema de isomorfismo de grafos (GIP por sus siglas en inglés) es computacionalmente difícil, cuya complejidad se encuentra estrictamente entre P y NP (a no ser que estas clases colapsen).

Otro problema relacionado es el de automorfismos de grafos.

PROBLEMA (Automorfismos de un grafo): *Determinar si el grupo de automorfismos de un grafo es trivial o no.*

Observar que los dos problemas presentados arriba busca respuestas del tipo *sí o no*, pero no busca encontrar elementos (isomorfismos o automorfismos) explícitamente.

Ejemplos 4.10. El grafo de la figura tiene grupo de automorfismo trivial.

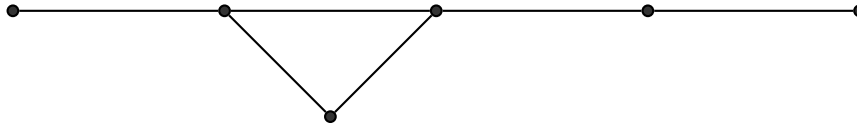


Figura 4.2: a.

Definición 4.11 (Matriz de adyacencia). Dado un grafo $G = (V, E)$, definimos su *matriz de adyacencia* como $A \in \text{Sym}(n)$

$$A_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \in E \\ 0 & \text{si no} \end{cases}$$

Comentario 4.12. Si A y B son matrices de adyacencia de dos grafos isomorfos, entonces $A = PBP^T$ para alguna matriz de permutación P .

En términos de las matrices de adyacencia, el problema de isomorfismo de grafos es: $\exists P \in \mathcal{P}(n) / A = PBP^T$?

Por otro lado, el denominado Graph Matching Problem (GMP)

PROBLEMA (Graph Matching Problem): *Encontrar el isomorfismo entre dos grafos (o el “más cercano” en algún sentido, si no son isomorfos). Esto es:*

$$P^* = \arg \min_{P \in \mathcal{P}(n)} \|A - PBP^T\|_F^2$$

El problema es que $\mathcal{P}(n)$ es un conjunto discreto con $n!$ elementos, por lo que computacionalmente no es viable probar todos los elementos para encontrar la solución.

Lema 4.13. Sea $\mathcal{N}(n) = \{M \in \mathbb{R}^{n \times n} : M_{ij} \geq 0\}$, entonces $\mathcal{P}(n) = \mathcal{O}(n) \cap \mathcal{N}(n)$.

El GMP es entonces $\arg \min_{P \in \mathcal{O}(n) \cup \mathcal{N}(n)} \|A - PBP^T\|_F^2$. Si nos olvidamos por el momento de la restricción de no negatividad, entonces queda formulado un problema en el grupo ortogonal.

Consideremos $f: \mathcal{O}(n) \rightarrow \mathbb{R}$, definida como $f(U) = \frac{1}{2}\|A - UBU^T\|_F^2$. Buscamos entonces minimizar f en $\mathcal{O}(n)$. Observemos que se tiene:

$$f(U) = \frac{1}{2}\|A\|_F^2 + \frac{1}{2}\|UBU^T\|_F^2 - \text{tr}(AUBU^T).$$

Luego, como los dos primeros términos son constantes para $U \in \mathcal{O}(n)$, se tiene

$$\arg \min_{U \in \mathcal{O}(n)} f(U) = \arg \min_{U \in \mathcal{O}(n)} -\text{tr}(AUBU^T).$$

Consideremos entonces $\tilde{f}(U) = -\text{tr}(AUBU^T)$ y calculemos el gradiente:

$$D\tilde{f}(U)\dot{U} = -\text{tr}(A\dot{U}BU^T) - \text{tr}(AUB\dot{U}^T).$$

Escribiendo $\dot{U} = U\Omega$, con $\Omega^T = -\Omega$ (de donde $\dot{U}^T = \Omega^T U^T = \Omega U^T$), se tiene

$$D\tilde{f}(U)U\Omega = \text{tr}(AUB\Omega U^T) - \text{tr}(AU\Omega BU^T) = \text{tr}([U^T AU, B]\Omega).$$

Como el gradiente es el que cumple $D\tilde{f}(U)U\Omega = \text{tr}((\text{grad } \tilde{f})^T U\Omega)$, resulta $\text{grad } \tilde{f}(U) = -U[U^T AU, B]$.

Por lo tanto, el flujo gradiente asociado a f en $\mathcal{O}(n)$ es $\dot{U} = U[U^T AU, B]$, que resulta completo por compacidad de $\mathcal{O}(n)$. Además, si las descomposiciones de A y B son $A = VD_1V^T$ y $B = WD_2W^T$, entonces se puede ver que los puntos críticos (que cumplen $[U^T AU, B] = 0$) son de la forma $P = W\Pi SV^T$, con $\Pi \in \mathcal{P}(n)$ y $S = \text{diag}(\pm 1, \dots, \pm 1)$. Hay por lo tanto $2^n n!$ puntos críticos, y el flujo converge a uno de ellos.

Observar que si A y B son matrices de adyacencia de grafos isomorfos, entonces la matriz de permutación P que conjuga A y B , es un punto crítico (y minimiza f en $\mathcal{O}(n)$).

Ahora, nada garantiza que el flujo tienda a una matriz de permutación. Consideremos entonces la función $g: \mathcal{O}(n) \rightarrow \mathbb{R}$, definida como $g(P) = \frac{2}{3} \text{tr}(P^T(P - P \circ P))$,

donde \circ es el producto de Hadarmard (el producto punto a punto de matrices).

Observemos primero que

$$g(P) = \frac{2}{3}n - \frac{2}{3} \operatorname{tr}(P^T(P \circ P)) = \frac{2}{3}n - \frac{2}{3} \sum_{ij} p_{ij}^3.$$

Por lo tanto, la función g se minimiza (sobre las matrices ortogonales) cuando $\sum_{ij} p_{ij}^3$ es lo más grande posible, y es fácil ver que esto sucede en las matrices de permutación. La idea será entonces construir un flujo, utilizando g , cuya solución sea convergente a una matriz de permutación, para después combinarlo con el flujo que estudiamos más arriba, que incluye la información de los grafos. No haremos las cuentas del gradiente ni de los puntos críticos, que se pueden encontrar en [Z].

El gradiente de g es $\operatorname{grad} g(P) = P((P \circ P)^T P - P^T(P \circ P))$, de donde el flujo gradiente asociado a f en $\mathcal{O}(n)$ es

$$\dot{P} = P(P^T(P \circ P) - (P \circ P)^T P).$$

Los puntos críticos son de la forma $P = S\Pi$, con Π matriz de permutación, y $S = \operatorname{diag}(\pm 1, \dots, \pm 1)$. De estos puntos críticos, las matrices de permutación son los únicos atractores del flujo.

Los autores de [Z] entonces combinan los dos flujos:

$$\dot{P} = (1-k)P[P^T A P, B] + k[P(P^T(P \circ P) - (P \circ P)^T P)],$$

pero poco se puede decir del comportamiento general de las soluciones.

Parte II

Métodos locales

Retracciones y búsqueda lineal

Cuando queremos optimizar una función $f: \mathbb{R}^n \rightarrow \mathbb{R}$ para encontrar un mínimo local, podemos considerar la siguiente estrategia iterativa,

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

donde x_k es dado, η_k es una *dirección* en que f decrece (por ejemplo $-\text{grad } f(x_k)$) y t_k es el *tamaño del paso*.

La idea de esta sección es estudiar una estrategia análoga para el caso en que f esté definida en una variedad M . La generalización natural en este caso es considerar $x_k \in M$ y buscar una dirección $\eta_k \in T_{x_k}M$ en la cual f disminuya su valor, para luego de algún modo "proyectar" el salto en M . Con este fin introduciremos la noción de *retracción*. Grosso modo, una retracción es una proyección del tangente T_xM en M . Tener una buena noción de retracción (y computacionalmente eficiente) nos permitiría aproximar líneas de flujo en variedades.

Definición 5.1 (Retracción). Una *retracción* en una variedad diferenciable M es una función diferenciable $R: TM \rightarrow M$ con las propiedades dadas a continuación. Denotando $R_x: T_xM \rightarrow M$, $R_x(\cdot) = R(x, \cdot)$, se tiene

1. $R_x(0) = x$, $0 \in T_xM$
2. Considerando la identificación canónica $T_0(T_xM) = T_xM$ se tiene que el mapa $DR_x(0): T_xM \rightarrow T_xM$ es la identidad en T_xM .

Observación 5.2. En algunas situaciones es posible que R esté localmente definido en un entorno de la sección 0 en TM .

Comentario 5.3. El segundo punto en la definición anterior puede entenderse del siguiente modo. Sea γ_ξ la curva en M dada por $\gamma_\xi(t) = R_x(t\xi)$, siendo $\xi \in T_xM$. Luego se tiene $\gamma_\xi(0) = x$, y

$$\dot{\gamma}_\xi(0) = \left. \frac{d}{dt} \right|_{t=0} \gamma_\xi(t) = \left. \frac{d}{dt} \right|_{t=0} R_x(t\xi) = DR_x(0)\xi = \xi,$$

i.e. $\dot{\gamma}_\xi(0) = \xi$. Por lo tanto, R_x proyecta rectas $t\xi$ por el origen en curvas en M que en 0 toman el valor x y su velocidad en $t = 0$ es ξ .

Observación 5.4. Otra aplicación importante de una retracción es linealizar localmente una función. Esto es, si $f : M \rightarrow \mathbb{R}$, luego, en un entorno de $x \in M$, podemos definir

$$\tilde{f}_x = f \circ R_x : T_xM \rightarrow \mathbb{R}.$$

En particular, $D\tilde{f}_x(0) = Df(x)$, y si M está dotada con estructura Riemanniana resulta $\text{grad}\tilde{f}_x(0) = \text{grad}f(x)$.

5.1. Ejemplos de retracciones

Hay muchas formas de definir retracciones en variedades. En particular toda variedad Riemanniana tiene una retracción natural dada por el *mapa exponencial* (ver Definición D.31). En los grupos de Lie por ejemplo, tenemos el mapa exponencial, que también verifica las propiedades de una retracción (ver Definición C.18). Sin embargo, con el objetivo de que la retracción nos ayude en problemas de optimización es necesario que el cálculo de la retracción sea computacionalmente eficiente, y este no es el caso de los ejemplos anteriores. Para el primer caso es necesario encontrar geodésietracc cual lo convierte en un problema equivalente al que queremos resolver. En el segundo, para el caso matricial, es necesario calcular series, funciones analíticas, lo cual no es un tema trivial. De hecho existe un área de estudio denominada *integración geométrica* donde se desarrollan técnicas particulares para abordar este problema en el caso de grupos de Lie (ver [IMKNZ]).

5.1.1. Retracciones en variedades encajadas

Supongamos que M es una variedad encajada en \mathbb{R}^n . En este caso $T_x M$ se puede identificar con el subespacio afín $x + \xi$ con $\xi \in T_x M$, donde abusando notación, identificamos $\xi \in T_x M$ con $\xi \in \mathbb{R}^n$. De esta manera nuestro objetivo es encontrar una manera de “proyectar” $x + t\xi \in \mathbb{R}^n$ en M . Veamos un resultado general que nos permitirá construir ejemplos.

Proposición 5.5. *Sea N una variedad (abstracta) tal que $\dim N = \text{codim } M$. Supongamos que existe un difeomorfismo $\phi : M \times N \rightarrow \text{Tub}(M) \subset \mathbb{R}^n$, donde $\text{Tub}(M)$ es un entorno abierto de M en \mathbb{R}^n (entorno tubular) tal que $\phi(x, e) = x \ \forall x \in M$ siendo $e \in N$. Luego el mapa $R_x(\xi) := \pi_1(\phi^{-1}(x + \xi))$, siendo $\pi_1 : M \times N \rightarrow M$ la proyección en la primer coordenada, es una retracción.*

Demostración. Que R es diferenciable surge de que $R(x, \xi) = \pi_1 \circ \phi^{-1}(x + \xi)$ es una composición de funciones diferenciables. Además como $\phi(x, e) = x$, resulta que $\phi^{-1}(x) = (x, e)$ y por lo tanto $R_x(0) = x$. Para probar que $DR_x(0)$ es la identidad, observar lo siguiente. Para t pequeño se tiene

$$DR_x(0)\xi = \left. \frac{d}{dt} \right|_{t=0} R_x(t\xi) = \left. \frac{d}{dt} \right|_{t=0} \pi_1 \circ \phi^{-1}(x + t\xi) = D\pi_1(x, e) \circ D\phi^{-1}(x)\xi.$$

Dado que $\phi(x, e) = x$ se tiene $D\phi(x, e)(\xi, 0) = \left. \frac{d}{du} \right|_{u=0} \phi(x + u\xi, 0) = \xi$, y por ser un difeomorfismo resulta $D\phi^{-1}(x)\xi = (\xi, 0)$. Luego el resultado sigue de que $D\pi_1(x, e)(\xi, 0) = \xi$. \square

Ejemplo 5.6 (Retracción en S^{n-1}). Sea $M = S^{n-1} \subset \mathbb{R}^n$ y sea $N = \{r \in \mathbb{R} : r > 0\} = \mathbb{R}_{>0}$ y sea $\phi : M \times \mathbb{R}_{>0} \rightarrow (\mathbb{R}^n)^*$ dada por $\phi(x, r) = rx$. ϕ es un difeo donde $\phi^{-1}(x + \xi) = \frac{x + \xi}{\|x + \xi\|}$ y $\phi(x, 1) = x$ para todo $x \in S^{n-1}$. Luego la retracción inducida es

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

Observar que $R_x(\xi) = \text{argmin}_{y \in S^{n-1}} \|x + \xi - y\|$.

Ejemplo 5.7 (Retracciones en $\mathcal{O}(n)$). (i): Sea $M = \mathcal{O}(n) \subset \mathbb{R}^{n \times n}$. La descomposición QR de una matriz nos brinda un ejemplo de retracción a $\mathcal{O}(n)$.

Sea $\phi : \mathcal{O}(n) \times \mathcal{R}_+(n) \rightarrow \text{Gl}(n, \mathbb{R})$ dada por $\phi(Q, R) = QR$, siendo $\mathcal{R}_+(n)$

matrices triangulares superiores con elementos > 0 en la diagonal. Se puede probar que ϕ es un difeo. Sea $qf = \pi_1 \circ \phi^{-1}$ el mapa que envía a una matriz a su factor ortogonal Q de la descomposición QR . Luego

$$R_U(U\Omega) = qf(U + U\Omega) = qf(U(\text{Id}_n + \Omega)) = Uqf(\text{Id}_n + \Omega)$$

es una retracción.

(ii) Descomposición Polar: Toda matriz A en $\mathbb{G}l(n, \mathbb{R})$ puede escribirse como $A = QP$ con Q ortogonal y P simétrica definida positiva. Luego $\phi : \mathcal{O}(n) \times \text{Sym}_+(n) \rightarrow \mathbb{G}l(n, \mathbb{R})$ induce la retracción

$$R_U(U\Omega) = U(\text{Id} + \Omega)(\text{Id} - \Omega^2)^{-1/2}.$$

(iii) Cayley Transform:

$$R_U(U\Omega) = U(\text{Id} - \frac{1}{2}\Omega)^{-1}(\text{Id} + \frac{1}{2}\Omega)$$

(iv) Mapa exponencial:

$$R_U(U\Omega) = U \exp(\Omega),$$

donde \exp es el mapa exponencial de matrices. Observar que esta retracción puede servir en cualquier grupo de Lie.

Ejemplo 5.8 (Retracción en $St(k, n)$). La descomposición QR también induce una retracción en $St(k, n)$, a saber

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

siendo $qf(A)$ el factor $Q \in St(k, n)$ de $A = QR$ con $A \in \mathbb{R}^{n \times k}$ de rango máximo y $R \in \mathcal{R}_+(n)$.

5.2. Retracciones en variedades cociente

Supongamos que tenemos una variedad cociente $\tilde{M} = M / \sim$ donde \sim relación de equivalencia y sea $\pi : M \rightarrow \tilde{M}$ la proyección canónica (todos los casos que estudiaremos son espacios homogéneos). Nuestra idea es definir una retracción en el

cociente a partir de una retracción en el ambiente. Sea $R : TM \rightarrow M$ una retracción en M . Una forma natural de definir una retracción en \tilde{M} es la siguiente:

Sea $\xi \in T_{[x]}\tilde{M}$, donde $\xi_x \in T_xM$ satisface $D\pi(x)\xi_x = \xi$. Luego definimos $\tilde{R}_{[x]}(\xi) = \pi(R_x(\xi_x))$. Para que esta definición tenga sentido es necesario pedir que $\pi(R_x(\xi_x)) = \pi(R_{x'}(\xi_{x'}))$ siendo $x \sim x'$ y $\xi_x \sim \xi_{x'}$, es decir $D\pi(x)\xi_x = D\pi(x')\xi_{x'}$.

Observación 5.9. En una variedad cociente siempre se puede definir una *distribución horizontal* que es una forma de identificar de manera única un representante del tangente $T_{[x]}\tilde{M}$ en cada uno de los representantes de $[x]$. Esto es, como $\pi^{-1}([x])$ es una variedad, $v_x := T_x\pi^{-1}([x])$ es su espacio tangente, y por lo tanto podemos definir un espacio complementario H_x en T_xM , es decir $T_xM = v_x \oplus H_x$. Una distribución horizontal es un mapa que asigna a cada $x \in M$ un espacio horizontal.

Observación 5.10. En el caso de espacios homogéneos Riemannianos, lo anterior se hace de manera natural tomando el ortogonal al núcleo: $H_x = (\text{Ker } F\pi(x))^\perp$. Luego existe un único $\xi_x \in H_x$ tal que $D\pi(x)\xi_x = \xi$.

Proposición 5.11. Si $\tilde{M} = M / \sim$ con distribución horizontal. Sea R retracción en M tal que $\pi(R_x(\xi_x)) = \pi(R_{x'}(\xi_{x'}))$ para todo $x \sim x'$ y $\xi_x \in H_x$, $\xi_{x'} \in H_{x'}$, entonces

$$\tilde{R}_{[x]}(\xi) := \pi(R_x(\xi_x))$$

es una retracción en \tilde{M} , siendo $\xi = D\pi(x)\xi_x$.

Demostración. La condición que le pedimos a R hace que \tilde{R} esté bien definida. Además $\tilde{R}_{[x]}(0) = \pi(R_x(0)) = [x]$. La condición de rigidez resulta

$$D\tilde{R}_{[x]}(0)\eta = D\pi(x)DR_x(0)\eta_x = D\pi(x)\eta_x = \eta$$

para todo $\eta \in T_{[x]}\tilde{M}$. □

Ejemplo 5.12. (i) (Retracciones en $\mathbb{P}(\mathbb{R}^n)$.) El espacio $\mathbb{P}(\mathbb{R}^n) = \mathbb{R}^n \setminus \{0\} / \mathbb{R}^*$ lo podemos ver como espacio homogéneo de la acción del grupo de Lie \mathbb{R}^* en $\mathbb{R}^n \setminus \{0\}$. Luego tomando la retracción trivial $R_x(\dot{x}) = x + \dot{x}$, se tiene que $\tilde{R}_{[x]}([\dot{x}]) = \pi(x + \dot{x})$. Una distribución horizontal viene dada por

$$H_x := \{\dot{x} \in \mathbb{R}^n : \dot{x}^T x = 0\},$$

si $x' = \alpha x$, $\dot{x}' = \alpha \dot{x}$, satisfacen $\pi(\alpha x + \alpha \dot{x}) = \pi(x + \dot{x})$ y por lo tanto estamos en las hipótesis de la proposición.

(ii) (Retracción en $\text{Grass}(k, n)$.) La grasmaniana la podemos obtener como $\text{Grass}(k, n) = \text{St}(k, n)/\mathcal{O}(k)$. Sea R la retracción en $\text{St}(k, n)$ dada por $qf(-)$. Es un ejercicio verificar que cumple las hipótesis de la proposición. Luego, $R_{(X)}(\xi) := \langle qf(X + \xi) \rangle$ define una retracción.

5.3. Algoritmos de búsqueda lineal

En el contexto de variedades tenemos que los *algoritmos de búsqueda lineal* son de la forma $x_{k+1} = R_{x_k}(t_k \eta_k)$, donde $t_k \in \mathbb{R}$, $\eta_k \in T_{x_k}M$. Ya tenemos definido R_{x_k} , ahora nos concentraremos en determinar t_k y η_k .

Definición 5.13 (Sucesión gradiente relacionado). Sea $f : M \rightarrow \mathbb{R}$ función costo en M variedad Riemanniana, y $\{x_k\} \subset M$. Una sucesión $\{\eta_k\}$ en TM con $\eta_k \in T_{x_k}M$ se dice *sucesión gradiente relacionado* si para toda subsucesión $\{x_{k'}\}$, la correspondiente $\{\eta_{k'}\}$ satisface que es acotada

$$\lim_{k' \rightarrow \infty} \langle \text{grad} f(x_{k'}), \eta_{k'} \rangle < 0.$$

Definición 5.14 (Punto de Armijo). Sea $f : M \rightarrow \mathbb{R}$ función costo en una variedad Riemanniana M , con retracción R . Sea $x \in M$, $\eta \in T_x M$, $\bar{\alpha} > 0$, $\beta, \sigma \in (0, 1)$. El *punto de Armijo* es $\eta^A = t^A \eta$ donde $t^A = \beta^m \bar{\alpha}$ siendo m el menor natural tal que

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

El escalar t^A se llama *paso de Armijo*.

La idea intuitiva para la elección del paso de Armijo es la siguiente. Estamos parados en un punto $x \in M$, tenemos elegida una dirección en el tangente sobre la cual nos queremos mover (por ejemplo la dirección opuesta al gradiente), y queremos elegir cuánto nos movemos en esa dirección.

La regla de Armijo lo elige de la siguiente forma. Empezamos con un paso $\bar{\alpha}$, nos movemos desde x en la dirección η un paso $\bar{\alpha}$ y evaluamos la función en ese

punto. Si nos satisface la ganancia (es decir, cuánto valía f en x , y cuánto vale en el nuevo punto), nos quedamos con ese paso, y sino, achicamos el paso por un factor $\beta < 1$, y volvemos a intentar. Repetimos este procedimiento (reduciendo el paso por β cada vez), hasta que se cumpla la condición expresada en la desigualdad. Esta condición dice, básicamente, que esperamos una reducción en el valor funcional de al menos una fracción σ de lo que se reduciría si la función fuese lineal (esto es, el producto interno del gradiente con el vector $\beta^m \bar{\alpha} \eta$).

ALGORITMO: ARMIJO MODIFICADO (ACELERADO)

- *Requerimientos* $\bar{\alpha} > 0, c, \beta, \sigma \in (0, 1)$.
 - **Input:** $x_0 \in M$
 - **Output:** sucesión $\{x_k\}$
- >> **for** $k = 0, 1, 2, \dots$
- >> elegir η_k tal que $\{\eta_k\}$ es gradiente relacionado
- >> elegir x_{k+1} tal que
- $$f(x_k) - f(x_{k+1}) \geq c(f(x_k) - f(R_{x_k}(t_k^A \eta_k))) (**)$$
- donde t_k^A es el paso de Armijo para $\bar{\alpha}, \beta, \sigma, \eta_k$.
- >> **End for.**

Observación 5.15. La condición $(**)$ se verifica para $x_{k+1} = R_{x_k}(t_k^A \eta_k)$. Por lo tanto este algoritmo permite otras posibles elecciones.

Comentario 5.16. Se puede probar que si $\{x_k\}$ es el output del algoritmo, entonces todo punto de acumulación de $\{x_k\}$ es punto crítico de f .

Ejemplo 5.17. (Cociente de Rayleigh) $f: S^{n-1} \rightarrow \mathbb{R}, f(x) = x^T A x$, con $A \in \text{Sym}(n)$. Utilizando la retracción $R_x(\xi) = \frac{x+\xi}{\|x+\xi\|}$ se tiene que $\eta_k = -2(Ax_k - x_k x_k^T A x_k)$ y encontramos el mínimo $m \geq 0$ tal que $f(R_{x_k}(\bar{\alpha} \beta^m \eta_k)) \leq f(x_k) - \sigma \bar{\alpha} \beta^m \eta_k^T \eta_k$ y tomamos $x_{k+1} = R_{x_k}(\bar{\alpha} \beta^m \eta_k)$.

Observación 5.18. $f(R_{x_k}(t \eta_k))$ es una función racional cuadrática.

Geometría diferencial computacional

En este capítulo retomaremos nociones geométricas trabajadas a lo largo del curso pero desde un punto de vista que sea de mayor utilidad en el álgebra lineal numérica. En particular, revisaremos la variedad de Stiefel con dos miradas distintas -como subvariedad euclídea y como variedad cociente- y daremos expresiones para las geodésicas en ambos casos. (Ver Sección 2.3.1 y el ejemplo 6 en C.39.)

(El material de este capítulo fue extraído de Edelman et al. [EAS].)

El primer cambio de enfoque resulta de cómo observar las variedades. En general, en geometría diferencial se utiliza las coordenadas para parametrizar una variedad. Por ejemplo, la esfera se puede parametrizar tomando coordenadas esféricas,

$$x = \cos u \sin v, \quad y = \sin u \sin v, \quad z = \cos v,$$

sin embargo, desde un punto de vista computacional es más eficiente utilizar coordenadas *extrínsecas* para describir la esfera:

$$(x, y, z) : \quad x^2 + y^2 + z^2 = 1.$$

Observar que estamos tomando más variables que las que se requieren, pero veremos que de esta manera encontraremos expresiones de la métrica, geodésicas, etc, con expresiones simples que pueden implementarse en una computadora.

6.1. Variedad de Stiefel euclídea

Recordar que la variedad de Stiefel, como subvariedad de $\mathbb{R}^{n \times p}$, está dada por

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

Diferenciando el producto $Y^T Y = \text{Id}_p$, obtenemos que el tangente está dado por

$$T_Y St(n, p) = \{Y^T \Delta \in \text{Skew}(p)\}.$$

(Cf. Proposición 2.40.)

Utilizando la métrica ambiente, del Lema 2.44, se tiene que el espacio normal está dado por

$$N_Y St(n, p) = \{YN : N^T = N \in \mathbb{R}^{p \times p}\}.$$

Además, las proyecciones ortogonales respectivas están dadas por

$$\pi_N(Z) = Y \text{Sym}(Z), \quad \pi_T(Z) = Y \text{Skew}(Z) + (\text{Id}_n - YY^T)Z,$$

siendo Sym y Skew las partes simétrica y antisimétrica. Toda matriz del tangente se puede escribir como

$$\Delta = YA + Y_\perp B \tag{6.1}$$

con $A \in \text{Skew}(p)$, $B \in \mathbb{R}^{n \times p}$ y Y_\perp tal que $YY^T + Y_\perp Y_\perp^T = \text{Id}_n$.

La *métrica euclídea* de $St(p, n)$ es simplemente la restricción al tangente de la métrica ambiente. Por lo tanto se define por

$$g_e(\Delta_1, \Delta_2) = \text{tr} \Delta_1^T \Delta_2.$$

siendo el lado derecho el producto interno de Frobenius $\mathbb{R}^{n \times p}$.

6.1.1. Geodésicas en $St(n, p)$

Las *geodésicas* son las curvas de “longitud mínima” entre dos puntos.

En el Capítulo D se definen las geodésicas como las curvas tales que la derivada covariante de la velocidad es nula (ver Definición D.26). En el caso euclidean

(i.e. variedades encajadas en algún \mathbb{R}^m) esto significa que aceleración normal a la superficie, siempre y cuando tengan la velocidad normalizada.

Principio Variacional

Estas dos propiedades de las geodésicas se relacionan a través de las ecuaciones de Euler-Lagrange como veremos a continuación.

Consideremos M una variedad encajada con la métrica Riemanniana inducida por el ambiente. Supongamos que tenemos una familia de curvas, parametrizadas de manera diferenciable por $s \in (-\varepsilon, \varepsilon)$, $\gamma_s(t)$, para $t \in [0, T]$, tales que todas comienzan y terminan en dos puntos fijos determinados. Esto es, $\gamma_s(0) = a$, y $\gamma_s(T) = b$, para todo s . (Esta familia de curvas puede pensarse como una variación-deformación diferenciable de γ_0 .)

Sea $F : \mathbb{R} \times TM \rightarrow \mathbb{R}$, y consideremos el funcional

$$I(s) = \int_0^T F(t, \gamma_s(t), \dot{\gamma}_s(t)) dt.$$

En el caso de $F(t, x, v) = \|v\|$, el funcional $I(s)$ resulta ser la longitud de la curva γ_s .

La idea es mostrar que los puntos críticos del funcional I son geodésicas.

$$\begin{aligned} \frac{d}{ds} I(s) &= \int_0^T \left\langle \frac{d}{ds} \gamma_s(t), \partial_x F \right\rangle + \left\langle \frac{d}{ds} \dot{\gamma}_s(t), \partial_v F \right\rangle dt \\ &= \int_0^T \left\langle \frac{d}{ds} \gamma_s(t), \partial_x F - \frac{d}{dt} \partial_v F \right\rangle + \left\langle \frac{d}{ds} \gamma_s(t), \frac{d}{dt} \partial_v F \right\rangle + \left\langle \frac{d}{ds} \dot{\gamma}_s(t), \partial_v F \right\rangle dt \end{aligned}$$

donde intercambiando el orden de derivación en γ_s , y usando integración por partes resulta:

$$\frac{d}{ds} I(s) = \left\langle \frac{d}{ds} \gamma_s(t), \partial_v F \right\rangle \Big|_0^T + \int_0^T \left\langle \frac{d}{ds} \gamma_s(t), \partial_x F - \frac{d}{dt} \partial_v F \right\rangle dt.$$

Cuando $s = 0$, resulta que el extremo izquierdo se hace igual cero por las condiciones en la familia de curvas, y por lo tanto resulta

$$\dot{I}(0) = \int_0^T \left\langle \frac{d}{ds} \gamma_0(t), \partial_x F - \frac{d}{dt} \partial_v F \right\rangle dt.$$

Observar que el mapa $t \mapsto \frac{d}{ds} \gamma_0(t)$ define un campo vectorial “transversal” a la velocidad de la curva γ_0 . (Se puede probar que dado un campo transversal a lo largo de γ_0 , siempre existe una variación de curvas con esa campo como campo trasversal.) Luego variando

dentro de todas las posibles deformaciones se concluye que $\gamma = \gamma_0$ es un punto crítico del funcional I si y sólo si

$$\partial_x F - \frac{d}{dt} \partial_v F \in T_{\gamma(t)} M^\perp. \quad (6.2)$$

En el caso de que $F(t, x, v) = \|v\|$, resulta que la condición (6.2) implica que

$$\dot{\gamma}(t) = \frac{d}{dt} \dot{\gamma}(t) \in T_{\gamma(t)} M^\perp.$$

Con esta información intentemos derivar la fórmula de la ecuación de una geodésica en $St(n, p)$.

Derivando dos veces la expresión $Y^T Y = \text{Id}$, obtenemos

$$\ddot{Y}^T Y + 2\dot{Y}^T \dot{Y} + Y^T \ddot{Y} = 0. \quad (6.3)$$

Por ser geodésica sabemos que $\ddot{Y} + YS = 0$ para S simétrica. Luego sustituyendo en (6.3), resulta que la ecuación de la geodésica está dada por

$$\ddot{Y} + Y(\dot{Y}^T \dot{Y}) = 0, \quad (6.4)$$

De la ecuación de la geodésica (6.4) se pueden obtener tres integrales de movimiento utilizando las siguientes definiciones:

$$C = Y^T Y, \quad A = Y^T \dot{Y}, \quad S = \dot{Y}^T \dot{Y}.$$

Luego resulta

$$\begin{cases} \dot{C} &= A + A^T \\ \dot{A} &= S - CS \\ \dot{S} &= [A, S] \end{cases}$$

De esta manera, si $Y(t)$ está en $St(n, p)$, \dot{Y} pertenece al tangente de $St(n, p)$, resulta que

$$\begin{cases} C(t) &= \text{Id} \\ A(t) &= A(0) \text{ es antisimétrica} \\ S(t) &= e^{At} S(0) e^{-At}. \end{cases}$$

(Comparar la última ecuación con la prueba de Lema 2.67.)

De esta manera, con las integrales de movimiento, se puede obtener una expresión muy sencilla de las geodésicas.

Observando que

$$\frac{d}{dt}(Ye^{At}, \dot{Y}e^{At}) = (Ye^{At}, \dot{Y}e^{At}) \begin{pmatrix} A & -S(0) \\ \text{Id} & A \end{pmatrix},$$

luego resulta

$$Y(t) = (Y(0), \dot{Y}(0)) \exp\left(t \begin{pmatrix} A & -S \\ \text{Id} & A \end{pmatrix}\right) \text{Id} e^{-At}, \quad (6.5)$$

donde $A = A(0) = Y(0)^T \dot{Y}(0)$. La expresión (6.5) es una fórmula cerrada para las geodésicas en la variedad de Stiefel con la métrica inducida por el ambiente.

En el caso $p = n$, la variedad de Stiefel coincide con el grupo ortogonal $\mathcal{O}(n)$, y una fórmula cerrada para la ecuación de la geodésicas puede obtenerse simplemente observando que

$$A = Q(t)^T \dot{Q}(t) = Q(0)^T \dot{Q}(0),$$

de donde se concluye que

$$Q(t) = Q(0)e^{At}.$$

◇ **6.1.** En particular se tiene que la distancia entre dos matrices ortogonales Q_1 y Q_2 (en una misma componente) está dada por

$$d(Q_1, Q_2) = \sqrt{\sum \theta_k^2},$$

donde $\{e^{i\theta_k}\}$ son los valores propios de $Q_1^T Q_2$.

En general las ecuaciones de las geodésicas están dadas en coordenadas locales utilizando las fórmulas de Christoffel (ver Capítulo D).

6.2. Variedad de Stiefel cociente

Hemos estudiado la geometría de la variedad de Stiefel, y en particular hemos obtenido expresiones cerradas de las geodésicas en el grupo ortogonal. Esto nos

permitirá estudiar la geometría de variedades cocientes del grupo ortogonal de manera sencilla.

En lo que sigue estudiaremos la variedad de Stiefel como cociente del grupo ortogonal. Es fácil ver que

$$St(n, p) = \mathcal{O}(n) / \mathcal{O}(n-p)$$

(ver Comentario 2.38 y ejemplo 6 en C.39).

Identificaremos a las clases por $[Q] \in St(n, p)$, donde la clase está determinada por el conjunto de matrices

$$[Q] = \left\{ Q \begin{pmatrix} \text{Id}_p & 0 \\ 0 & Q_{n-p} \end{pmatrix} : Q_{n-p} \in \mathcal{O}(n-p) \right\} \quad (6.6)$$

(Observar que en el caso euclideo, la clase $[Q]$ se identifica con la matriz $Q \text{Id}_{n,p}$, i.e. con las primera p columnas de Q).

Dotando el grupo ortogonal con la métrica Riemanniana inducida por el ambiente, podemos definir los siguientes espacios.

El *espacio vertical* $\Phi \subset T_Q \mathcal{O}(n)$ es el espacio tangente a la fibra, y su ortogonal es el *espacio horizontal* $\Delta = \Phi^\perp \subset T_Q \mathcal{O}(n)$.

Lema 6.2. *El espacio vertical y el espacio horizontal en Q , están dados por*

$$\Phi = \left\{ Q \begin{pmatrix} 0 & 0 \\ 0 & C \end{pmatrix} : C \in \text{Skew}(n-p) \right\}$$

y

$$\Delta = \left\{ Q \begin{pmatrix} A & -B^T \\ B & 0 \end{pmatrix} : A \in \text{Skew}(p) \right\}.$$

◇ **6.3.** Probar el resultado anterior.

Como vimos en la Sección C.4 sobre espacios homogéneos, el espacio horizontal se identifica con el tangente a la variedad cociente. Sobre esta podemos poner la métrica inducida por el grupo ortogonal (escalada por $1/2$), i.e. si $\Delta_1, \Delta_2 \in \Delta$ se tiene

$$\begin{aligned} g_c(\Delta_1, \Delta_2) &= \frac{1}{2} \text{tr} \left(\left[Q \begin{pmatrix} A_1 & -B_1^T \\ B_1 & 0 \end{pmatrix} \right]^T Q \begin{pmatrix} A_2 & -B_2^T \\ B_2 & 0 \end{pmatrix} \right) \\ &= \frac{1}{2} \text{tr}(A_1^T A_2) + \text{tr}(B_1^T B_2). \end{aligned}$$

A esta métrica le llamamos *métrica canónica* o *homogénea* (para no confundirla con la euclídea).

Es importante destacar que la variedad de Stiefel con esta métrica es diferente al caso euclideo, (salvo en los casos extremos $p = 1$ o n).

◇ **6.4.** Mostrar con un ejemplo que las métricas pueden no coincidir.

◇ **6.5.** Probar que si $\Delta \in \mathbb{R}^{n \times p}$, satisface $Y^T \Delta = 0$, entonces la “métrica canónica” en este vector tangente satisface

$$g_c(\Delta, \Delta) = \text{tr}(\Delta^T (\text{Id} - YY^T) \Delta).$$

Observar que la curva en $\mathcal{O}(n)$,

$$Q(t) = Q(0) \exp t \begin{pmatrix} A & -B^T \\ B & 0 \end{pmatrix}, \quad (6.7)$$

es una geodésica en el grupo ortogonal, y además su derivada

$$\dot{Q}(t) = Q(t) \begin{pmatrix} A & -B^T \\ B & 0 \end{pmatrix},$$

satisface que pertenece al espacio horizontal. Utilizando la propiedad de que las geodésicas minimizan distancia se concluye que estas curvas también minimizan en el cociente y por lo tanto son geodésicas también en la variedad de Stiefel vista como espacio homogéneo.

Por lo tanto hemos concluido

$$\text{geodésicas en } St(k, n) = [Q(t)],$$

donde $[Q(t)]$ está dado en (6.6) y (6.10).

6.3. Geometría de Grass(p, n)

El caso de la grassmaniana Grass(p, n) es muy similar a la sección anterior, donde se puede definir como el cociente $\mathcal{O}(n)/\mathcal{O}(p) \times \mathcal{O}(n-p)$. La clase de equivalencia de $[Q]$ es el conjunto de matrices tales que las primeras p columnas

generan el mismo subespacio vectorial de dimensión p en \mathbb{R}^n . De esta manera $\text{Grass}(p, n)$ es el conjunto de clases de equivalencia

$$[Q] = \left\{ Q \begin{pmatrix} Q_p & 0 \\ 0 & Q_{n-p} \end{pmatrix} : Q_p \in \mathcal{O}(p), Q_{n-p} \in \mathcal{O}(n-p) \right\} \quad (6.8)$$

◇ **6.6.** Análogamente a lo anterior el *espacio vertical* $\Phi \subset T_Q \mathcal{O}(n)$ (es el espacio tangente a la fibra), y su ortogonal es el *espacio horizontal* $\Delta = \Phi^\perp \subset T_Q \mathcal{O}(n)$, están dados por

$$\Phi = \left\{ Q \begin{pmatrix} A & 0 \\ 0 & C \end{pmatrix} : A \in \text{Skew}(p), C \in \text{Skew}(n-p) \right\}$$

y

$$\Delta = \left\{ Q \begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix} : B \in \mathbb{R}^{(n-p) \times p} \right\}.$$

La *métrica canónica en* $\text{Grass}(p, n)$ es la restricción al espacio horizontal (escalada por un factor de $1/2$) de la métrica canónica en el grupo ortogonal. Por lo tanto se tiene

$$g_c(\Delta_1, \Delta_2) = \text{tr}(B_1^T B_2) \quad (6.9)$$

(donde la notación sigue la notación de los espacios.)

◇ **6.7.** Observar que en este caso, la métrica canónica en $\text{Grass}(p, n)$ sí coincide con la métrica euclídea.

La geodésica en $\mathcal{O}(n)$,

$$Q(t) = Q(0) \exp t \begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix}, \quad (6.10)$$

tiene velocidad en el espacio horizontal

$$\dot{Q}(t) = Q(t) \exp t \begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix},$$

en cada punto $Q(t)$, y por lo tanto podemos concluir

$$\text{geodésicas en } \text{Grass}(p, n) = [Q(t)],$$

donde $[Q(t)]$ está dado en (6.8) y (6.9).

6.3.1. Grass(p, n) como cociente de $St(p, n)$

En esta sección, motivado por el interés computacional, estudiaremos la grassmaniana como espacio homogéneo de la variedad de Stiefel, a saber, $\text{Grass}(p, n) = St(p, n)/\mathcal{O}(p)$.

De esta manera, una matriz $Y \in St(p, n)$ representará toda la clase de equivalencia dada por

$$[Y] = \{YQ_p : Q_p \in \mathcal{O}(p)\}, \quad (6.11)$$

es decir el subespacio generado por las columnas de Y .

Recordando la expresión (6.1) de vectores tangentes a $St(p, n)$ en Y , se tiene $\Delta = YA + Y_\perp B$ con $A \in \text{Skew}(p)$, $B \in \mathbb{R}^{(n-p) \times p}$ y Y_\perp tal que $YY^T + Y_\perp Y_\perp^T = \text{Id}_n$.

◇ **6.8.** De (6.11) se concluye que el espacio vertical y el espacio horizontal están dados por

$$\Phi = \{YA : A \in \text{Skew}(p)\}, \quad \Delta = \{Y_\perp B : B \in \mathbb{R}^{n \times p}\}, \quad (6.12)$$

respectivamente. En particular, el espacio tangente se identifica con el subespacio de matrices $\dot{X} \in \mathbb{R}^{n \times p}$ tales que

$$Y^T \dot{X} = 0.$$

Dar una interpretación geométrica a esta ecuación.

En el siguiente resultado (sin demostración) daremos una expresión alternativa para las geodésicas en $\text{Grass}(p, n)$.

Teorema 6.9 (Geodésicas en $\text{Grass}(p, n)$). *Sea $Y(t) = Q \exp t \begin{pmatrix} 0 & -B^T \\ B & 0 \end{pmatrix} \text{Id}_{n,p}$, con $Y(0) = Y$, $\dot{Y}(0) = H$, entonces*

$$Y(t) = (YV \quad U) \begin{pmatrix} \cos \Sigma t \\ \sin \Sigma t \end{pmatrix} V^T,$$

donde $H = U \Sigma V^T$ es su descomposición en valores singulares (reducida). □

Gradiente de funciones en $\text{Grass}(p, n)$

En las sección anterior discutimos sobre las geodésicas en $\text{Grass}(p, n)$. En esta sección se describirá el gradiente de una función definida en $\text{Grass}(p, n)$.

Sea $f : \text{Grass}(p, n) \rightarrow \mathbb{R}$ una función diferenciable. El gradiente de f en $[Y]$ es definido por la proyección ortogonal del gradiente ambiente en el espacio horizontal. Es fácil ver que

$$\text{grad } f([Y]) = f_Y - Y^T Y f_Y, \quad (6.13)$$

siendo $f_Y = ((\frac{\partial f}{\partial Y_{ij}}))$ la matriz $n \times p$ que representa el gradiente ambiente.

◇ **6.10.** Dar una nueva fórmula de la fórmula del gradiente derivando f a lo largo de una geodésica.

6.4. Hessiano de funciones

Como veremos en el próximo capítulo, el método de Newton requiere del hessiano de una función definida en una variedad. (En la Sección D.7 se define el hessiano de una función definida en variedad con una conexión.)

Sea $f : M \rightarrow \mathbb{R}$ una función diferenciable en una variedad Riemanniana M . Entonces el hessiano de f en p está dado por la fórmula

$$\langle \text{Hess}_p f(u), u \rangle = \left. \frac{d^2}{dt^2} f \circ \gamma(t) \right|_{t=0},$$

siendo $\gamma(t)$ una geodésica tal que $\gamma(0) = p$ y $\dot{\gamma}(0) = u$.

A modo de ejemplo daremos expresiones del hessiano de funciones definidas en la variedad de Stiefel y en la grassmaniana.

6.4.1. Hessiano de una función en $St(p, n)$ y $\text{Grass}(p, n)$

◇ **6.11.** Sea $f : St(p, n) \rightarrow \mathbb{R}$ una función C^∞ . Entonces se tiene

$$\begin{aligned} & \text{Hess}_p f(\Delta_1, \Delta_2) \\ &= f_{YY}(\Delta_1, \Delta_2) + \frac{1}{2} \text{tr}((f_Y^T \Delta_1 Y^T + Y^T \Delta_1 f_Y^T) \Delta_2) - \frac{1}{2} \text{tr}((Y^T f_Y + f_Y^T Y) \Delta_1^T \Pi \Delta_2), \end{aligned}$$

siendo $\Pi = \text{Id} - YY^T$, $f_Y = ((\frac{\partial f}{\partial Y_{ij}}))$ y $f_{YY}(\Delta_1, \Delta_2) = \sum_{i,j} \frac{\partial^2 f}{\partial Y_{ij} \partial Y_{kl}} (\Delta_1)_{i,j} (\Delta_2)_{k,l}$.

◇ **6.12.** Sea $f : \text{Grass}(p, n) \rightarrow \mathbb{R}$ una función C^∞ . Entonces se tiene

$$\text{Hess}_p f(\Delta_1, \Delta_2) = f_{YY}(\Delta_1, \Delta_2) - \text{tr}(\Delta_1^T \Delta_2 F_Y).$$

◇ **6.13.** Probar que si $\Delta = -(\text{Hess}_p f)^{-1} \xi$, entonces

$$f_{YY}(\Delta) - \Delta(Y^T f_Y) = -G.$$

Método de Newton

El *método de Newton* es el algoritmo más importante para encontrar ceros de una función.

7.1. Método de Newton en \mathbb{R}^n

El caso más simple es cuando tenemos una función $f: \mathbb{R} \rightarrow \mathbb{R}$, a la cual queremos encontrar x^* tal que $f(x^*) = 0$. El método de Newton, para un valor inicial x_0 , está dado por la sucesión dada por recurrencia

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad (k = 0, 1, 2, \dots)$$

La idea básica detrás del método de Newton es la siguiente. Tomando como punto de partida un punto $x_0 \in \mathbb{R}$ que no es un cero de f , aproximamos linealmente la función en ese punto y encontramos la raíz de la aproximación lineal.¹ Esto es, $x_1 = N_f(x_0)$ es la única solución de la ecuación

$$0 = f'(x_0)(x - x_0) + f(x_0).$$

¹Es claro que la única manera que el método falle es que la aproximación lineal sea una constante.

Otra forma de ver el método de Newton es considerar el sistema dinámico dado por el *operador de Newton*:

$$N_f : \mathbb{R} \rightarrow \mathbb{R}, \quad N_f(x) = x - \frac{f(x)}{f'(x)}.$$

Este operador está definido siempre que $f'(x) \neq 0$.

◇ **7.1.** Probar que si $x \in \mathbb{R}$ no es un punto crítico de f , entonces se tiene que $N_f(x) = x$ si y sólo si $f(x) = 0$.

Lo que dice el resultado anterior es que los ceros (no críticos) de f se corresponden con puntos fijos del operador de Newton N_f .

Una de las propiedades más importantes de la dinámica asociada al método de Newton es que los ceros no degenerados de f son puntos fijos *superatractores*. Esto es que $N'_f(x^*) = 0$.

◇ **7.2.** Probar que si $N_f(x^*) = x^*$ entonces $N'_f(x^*) = 0$. Luego observar que si x está cerca de x^* entonces $|N_f(x) - x^*| \leq C|x - x^*|^2$ para cierta constante C que puede ser elegida de manera uniforme en un entorno de x^* .

En el caso de $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, se procede de manera análoga. Dado $x_k \in \mathbb{R}^n$ se tiene que x_{k+1} es la solución del sistema lineal $f(x_k) + Df(x_k)(x - x_k) = 0$. Este sistema tiene solución (y única) siempre que $Df(x_k)$ sea invertible. De esta manera el método de Newton \mathbb{R}^n se define por

$$x_{k+1} = x_k - Df(x_k)^{-1} f(x_k). \quad (7.1)$$

Comentario 7.3. Otra forma de obtener el método de Newton es a través de la aproximación discreta de una ecuación diferencial en \mathbb{R}^n . Si consideramos el campo $V \in \mathfrak{X}(\mathbb{R}^n)$ dado por

$$V(x) = -Df(x)^{-1} f(x),$$

entonces la *aproximación de Euler* de la ecuación diferencial $\dot{x} = V(x)$, es el método de Newton. Esto es, si estamos parados en $x_k \in \mathbb{R}^n$, saltamos un paso en la dirección del campo $V(x_k)$:

$$x_{k+1} = x_k + V(x_k) = x_k - Df(x_k)^{-1} f(x_k).$$

7.2. Método de Newton en variedades

Supongamos que tenemos una función $f \in C^\infty(M)$, donde M es una variedad Riemanniana. Nuestro objetivo es encontrar los puntos críticos de f , i.e. $x \in M$ tal que $\text{grad } f(x) = 0$.

Si $M = \mathbb{R}^n$, basta usar el método de Newton para la función $\text{grad } f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, y por lo tanto dado $x_k \in \mathbb{R}^n$, el siguiente iterado satisface

$$\text{grad } f(x_k) + D(\text{grad } f)(x_k)(x_{k+1} - x_k) = 0.$$

El objetivo es extender lo anterior a una variedad no lineal.

Para eso hagamos una pequeña digresión sobre el método de Newton en \mathbb{R}^n . Sea F es un campo en \mathbb{R}^n , y supongamos que estamos parados en $x_k \in \mathbb{R}^n$. Linealizando el campo en x_k obtenemos una transformación lineal afín

$$x \mapsto F(x_k) + T(x - x_k),$$

donde $T = DF(x_k)$ es el operador derivada de F en x_k , y por lo tanto $T\eta$ es la derivada direccional de F en la dirección de η . (Observar que acá estamos identificando $T_{x_k}\mathbb{R}^n$ con \mathbb{R}^n .) Luego, notando por $\eta_k = x - x_k$, lo que estamos buscando es cuál es el vector η_k tal que $T(\eta_k) = -F(x_k)$.

Veamos el caso abstracto ahora. Supongamos que queremos encontrar ceros de un campo $\xi \in \mathfrak{X}(M)$ definido en una variedad M .

Para construir el método de Newton en la variedad se realizan los siguientes cambios.

- Linealizamos la variedad M en x_k , i.e. consideramos $T_{x_k}M$;
- tomamos la aproximación lineal del campo ξ en x_k . Para eso necesitamos una conexión ∇ que nos permita cuantificar cómo cambia el campo en distintas direcciones. La aproximación es el “jacobiano” $J(x_k) : T_{x_k}M \rightarrow T_{x_k}M$ dado por

$$J(x_k)\eta = \nabla_\eta \xi.$$

- buscamos $\eta_k \in T_{x_k}M$ tal que $J(x_k)\eta_k = -\xi_{x_k}$;

- y para finalizar proyectamos el vector η_k en M con alguna retracción definida a priori.

ALGORITMO: MÉTODO DE NEWTON PARA CAMPOS VECTORIALES

- *Requerimientos:* M variedad, R retracción, ∇ conexión afín, ξ campo en M .
- *Objetivos:* encontrar un cero de ξ .
- **Input:** primer iterado $x_0 \in M$
- **Output:** sucesión $\{x_k\}$ en M .

```
>> for  $k = 0, 1, \dots$ ,
    resolver la ecuación  $J(x_k)\eta_k = -\xi_{x_k}$  para la incognita  $\eta_k \in T_{x_k}M$ , donde
     $J(x_k)\eta_k = \nabla_{\eta_k}\xi$ ;
>>  $x_{k+1} := R(\eta_k)$ .
>> End for.
```

Observación 7.4. En el algoritmo anterior hay cierta libertad sobre cuál es la retracción a elegir. Como se discutió en el Capítulo 5, hay muchas variantes posibles y algunas más convenientes para el cálculo computacional. Ver el capítulo mencionado para una discusión sobre el tema.

Observación 7.5. Cuando M es una variedad Riemanniana, podemos hacer uso de lo estudiado en el Capítulo 5 sobre búsqueda lineal. En este caso podemos tomar la función penalización $f := \langle \xi, \xi \rangle$, la cual los ceros son precisamente los puntos buscados. Luego podemos combinar este algoritmo con el Algoritmo de Armijo en página 79. Esto es, se elige la dirección η_k como en Algoritmo de más arriba, para luego computar x_{k+1} para que satisfaga la condición de descenso

$$f(x_k) - f(x_{k+1}) \geq c(f(x_k) - f(R_{x_k}(t_k \eta_k))).$$

En el caso que ∇ es la conexión Riemanniana, entonces se tiene

$$D\langle \xi, \xi \rangle(x_k)(\eta_k) = \langle \nabla_{\eta_k} \xi, \xi \rangle + \langle \xi, \nabla_{\eta_k} \xi \rangle = -2\langle \xi, \xi \rangle_{x_k} < 0,$$

y por lo tanto si $\xi \neq 0$, se sigue que el vector η_k es una dirección descendente para f .

7.2.1. Método de Newton para puntos críticos

En particular tenemos el caso de $f : M \rightarrow \mathbb{R}$, $\xi = \text{grad } f$. En este caso el jacobiano es el Hessiano y por lo tanto hay que resolver

$$\text{Hess}_{x_k} f[\eta_k] = -\text{grad } f(x_k).$$

ALGORITMO: MÉTODO DE NEWTON PARA PUNTOS CRÍTICOS

- *Requerimientos:* M variedad Riemanniana, R retracción, ∇ conexión afín.
 - *Objetivo:* Encontrar puntos críticos de f .
 - **Input:** $x_0 \in M$.
 - **Output:** Sucesión $\{x_k\}$ en M .
- >> **for** $k = 0, 1, \dots$
 resolver la ecuación $\text{Hess}_{x_k} f[\eta_k] = -\text{grad } f(x_k)$, para la incógnita η_k .
- >> $x_{k+1} := R_{x_k}(\eta_k)$.
- >> **End for.**

Comentario 7.6. En general la solución $\text{Hess}_{x_k} f[\eta_k] = -\text{grad } f(x_k)$ no tiene porque estar en una dirección de descenso de f . De hecho

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

que no tiene porque ser negativa. Sí lo es en el caso de que el Hessiano sea definido positivo. En el caso de la conexión Riemanniana se puede probar que el Hessiano es definido positivo si y sólo si los valores propios son positivos.

7.2.2. Cociente de Rayleigh en la esfera

En esta sección veremos como el algoritmo de Newton geométrico nos permite convertir el cociente de Rayleigh en un algoritmo.

Para empezar recordemos el cociente de Rayleigh que estudiamos en la Sección [2.2](#).

Sea $f : S^{n-1} \rightarrow \mathbb{R}$, $f(x) = x^T Ax$, y sea $\tilde{f} : \mathbb{R}^n \rightarrow \mathbb{R}$ una extensión.

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

donde $P_x = \text{Id} - \frac{xx^T}{\|x\|^2}$, la proyección ortogonal sobre $T_x S^{n-1}$.

Tenemos la retracción $R_x(\xi) = \frac{x+\xi}{\|x+\xi\|}$ y la conexión afín $\nabla_\eta \xi = P_x(D\xi(x)[\eta])$.

Si $\xi = \text{grad } f$, entonces $\nabla_\eta \text{grad } f(x) = P_x(D \text{grad } f(x)[\eta]) = 2P_x(Ax - xx^T Ax) = 2(P_x A P_x \eta - \eta x^T Ax)$. La ecuación queda

$$\begin{cases} P_x A P_x \eta - \eta x^T Ax = -P_x(Ax) \\ x^T \eta = 0 \end{cases}$$

ALGORITMO: MÉTODO DE NEWTON PARA COCIENTE DE RAYLEIGH

▪ **Input:** $x_0 \in S^{n-1}$.

▪ **Output:** $\{x_k\}$

for $k = 0, 1, \dots$, resolver el sistema lineal

$$\begin{cases} P_{x_k} A P_{x_k} \eta_k - \eta_k x_k^T A x_k = -P_{x_k}(A x_k) \\ x_k^T \eta_k = 0 \end{cases}$$

para η_k

▪ $x_{k+1} := R_{x_k}(\eta_k)$.

7.2.3. Método de Newton en Grass(p, n)

Sea f una función C^∞ definida en Grass(p, n), i.e. $f(Y) = f(YQ)$ para todo $Q \in \mathcal{O}(p)$, donde $Y \in \mathbb{R}^{n \times p}$ tal que $Y^T Y = \text{Id}$.

Calculando las f_Y y f_{YY} según la Sección 6.4.1, podemos utilizar el método de Newton para minimizar f .

ALGORITMO: MÉTODO DE NEWTON MINIMIZACIÓN SOBRE $\text{Grass}(p, n)$

- **Input:** $Y_0 \in \mathbb{R}^{n \times p}$, tal que $Y_0^T Y_0 = \text{Id}$;
- >> for $k = 0, 1, \dots$,
 Calculamos gradiente: $G_k = f_{Y_k} - Y_k Y_k^T f_{Y_k}$
- >> Calculamos salto $\Delta_k = -\text{Hess}^{-1} G_k$:
 Δ_k tal que $Y_k^T \Delta_k = 0$, y $f_{Y_k Y_k}(\Delta_k) - \Delta(Y_k^T f_{Y_k}) = -G$
- >> $Y_{k+1} = R_{Y_k}(\Delta_k)$

Del Teorema 6.9 se puede cambiar la retracción por un cálculo explícito de la geodésica (siempre que conozcamos la descomposición en valores singulares de Y_0).

Parte III

Apéndice

Topología

En este capítulo detallaremos las principales definiciones y resultados sobre *espacios topológicos* que utilizamos en estas notas

A.1. Relaciones de equivalencia.

Definiciones A.1. ■ Una *relación* en X es un subconjunto del producto cartesiano $X \times X$. Dada una relación $\mathcal{R} \subset X \times X$ escribiremos $x\mathcal{R}y$ cuando $(x, y) \in \mathcal{R}$.

- Una *relación de equivalencia* en X es una relación \mathcal{R} que cumple con las siguientes propiedades:
 - i. $x\mathcal{R}x$ para todo $x \in X$ (reflexiva)
 - ii. $x\mathcal{R}y$ implica $y\mathcal{R}x$ (simétrica)
 - iii. $x\mathcal{R}y$ y $y\mathcal{R}z$ implica $x\mathcal{R}z$ (transitiva)

Para denotar las relaciones de equivalencia usaremos el símbolo \sim .

Definiciones A.2. Consideremos una relación de equivalencia \sim en X .

- La *clase de equivalencia* de un elemento $x_0 \in X$, se define como el subconjunto de X formado por

$$[x_0] := \{x \in X : x \sim x_0\}.$$

- Al conjunto formado por las clases de equivalencia $\{[x] : x \in X\}$ se denomina *espacio cociente* o *conjunto cociente*.
- A la función $\pi : X \rightarrow X/\sim$ que a $x \mapsto [x]$ se le llama *proyección al cociente*

A.2. Espacios topológicos

Definición A.3 (Topología). Sea X un conjunto. Una familia τ de subconjuntos de X es una *topología* de X si se cumplen:

1. El conjunto vacío y el conjunto X pertenecen a τ .
2. Si A_1, \dots, A_n pertenecen a τ , entonces $\bigcap_1^n A_j$ pertenece a τ .
3. Si I es un conjunto y A_α es un subconjunto de X que pertenece a τ para cada $\alpha \in I$, entonces $\bigcup_{\alpha \in I} A_\alpha$ pertenece a τ .

Si τ es una topología en X , decimos que el par (X, τ) es un *espacio topológico*. Los elementos de τ se ll

Definiciones A.4. Sea (X, τ) un espacio topológico.

- Un conjunto $B \subset X$ se dice *cerrado* si su complemento es abierto. Por lo tanto se tiene que la unión de una cantidad finita de conjuntos cerrados es cerrada y la intersección de una cantidad arbitraria de cerrados es cerrada.
- Diremos que V es *entorno* de un punto x de X si se cumple que $x \in A \subset V$ para algún abierto A .
- Un punto $x \in X$ es de *acumulación* de un subconjunto A de X si todo entorno de x contiene algún punto de A diferente de x .

- Un punto $x \in X$ es *interior* de un subconjunto A de X si existe un entorno V de x que está contenido en A .
- Un punto $x \in X$ es *frontera o borde* de un subconjunto A de X si todo entorno V de x intersecta a A y a A^c .
- Una sucesión $\{x_n : n \in \mathbb{N}\}$ *converge* a un punto $x \in X$ si dado cualquier entorno V de x existe un n_0 tal que $x_n \in V$ para todo $n \geq n_0$.

aman los *abiertos* de la topología.

Definición A.5 (Espacio Hausdorff). Un espacio topológico es de *Hausdorff* (se llama también T_2) si se cumple que para todo par de puntos distintos x e y , existen entornos V de x y W de y tales que $V \cap W = \emptyset$.

A.3. Funciones continuas y homeomorfismos

Definición A.6 (Función continua). Una función $f : X \rightarrow Y$ entre espacios topológicos (X, τ_X) e (Y, τ_Y) es *continua* si $f^{-1}(A) \in \tau_X$ para cada $A \in \tau_Y$.

Obviamente la continuidad depende de las topologías en X e Y . Por ejemplo, si τ_1 y τ_2 son topologías en un conjunto X , entonces la función identidad $id : (X, \tau_1) \rightarrow (X, \tau_2)$ es continua si y sólo si $\tau_1 \supset \tau_2$, en este caso diremos que la topología τ_1 es *más fina* que τ_2 . Por eso, cuando hay riesgo de confusión, escribimos $f : (X, \tau_X) \rightarrow (Y, \tau_Y)$, para dar a entender con qué topologías estamos considerando dominio y codominio.

Definición A.7 (Continuidad en un punto). Sean (X, τ_X) e (Y, τ_Y) espacios topológicos. Una función $f : X \rightarrow Y$ es *continua en x* si para todo entorno V de $f(x)$ existe un entorno U de x tal que $f(U) \subset V$.

Veamos algunas equivalencias:

◇ **A.8.** Sea $f : (X, \tau_X) \rightarrow (Y, \tau_Y)$. Son equivalentes:

1. f es continua.

2. f es continua en x para todo $x \in X$.

Definición A.9 (Homeomorfismo). Una función biyectiva entre espacios topológicos se dice un *homeomorfismo* si tanto f como f^{-1} son continuas. En este caso, decimos que los espacios son *homeomorfos*.

A.4. Bases

En esta sección veremos cómo generar topologías a partir de otras conocidas.

Observar que cuando definimos una topología en \mathbb{R} lo hicimos declarando que los abiertos son aquellos conjuntos A tales que para todo $x \in A$ existe un intervalo (a, b) tal que $x \in (a, b) \subseteq A$. Análogamente, podríamos definir que los abiertos son los conjuntos formados por uniones de intervalos abiertos (¡verificar!). De esta manera vemos que para definir la topología antes mencionada en \mathbb{R} basta con considerar los intervalos de la forma $(a, b) \subset \mathbb{R}$. Al conjunto de estos intervalos se le llama *base* de la topología.

Definición A.10 (Base). Decimos que $\mathcal{B} \subset \mathcal{P}(X)$ es *base de una topología* de un conjunto X si se cumplen:

1. La unión de todos los elementos de \mathcal{B} es X .
2. Dados B_1 y B_2 elementos de \mathcal{B} y un punto $x \in B_1 \cap B_2$, existe $B \in \mathcal{B}$ tal que $x \in B \subset B_1 \cap B_2$.

Notar que si \mathcal{B} es una base de una topología de X y se define τ como el conjunto de todas las uniones de elementos de \mathcal{B} más el vacío, entonces τ es una topología de X , que se llama la *topología generada* por \mathcal{B} . La topología τ es la mínima que contiene a todos los elementos de \mathcal{B} . También se dice en este caso que \mathcal{B} es base de τ .

A.5. Espacio producto

Una manera natural de definir una topología en el producto cartesiano de dos espacios topológicos es la siguiente.

Definición A.11 (Topología producto). Sean (X, τ_X) , (Y, τ_Y) espacios topológicos. Sea $\mathcal{B}_{X \times Y}$ el conjunto de los productos $A \times B$ tales que $A \in \tau_X$ y $B \in \tau_Y$. Entonces $\mathcal{B}_{X \times Y}$ es base de una topología de $X \times Y$ que llamaremos la *topología producto* de los espacios X e Y . Observar que $\mathcal{B}_{X \times Y}$ no es una topología en general.

Es fácil ver que cuando las topologías τ_X e τ_Y están generadas por bases \mathcal{B}_X e \mathcal{B}_Y respectivamente, entonces la topología producto está generada por conjuntos de la forma $B \times B'$, donde $B \in \mathcal{B}_X$ y $B' \in \mathcal{B}_Y$.

Más en general, dados X_1, \dots, X_n espacios topológicos, podemos definir la topología producto en $X_1 \times \dots \times X_n$ como la generada por la colección $U_1 \times \dots \times U_n$, donde U_i es abierto de X_i .

El espacio producto $X \times Y$ tiene dos mapas naturales $\pi_1 : X \times Y \rightarrow X$, $\pi_2 : X \times Y \rightarrow Y$, llamadas *proyecciones*, definidas por $\pi_1(x, y) = x$, $\pi_2(x, y) = y$.

◇ **A.12.** Las proyecciones son continuas.

◇ **A.13.** La topología producto es la topología más menos fina que hace a las proyecciones continuas.

Definición A.14 (Fibra). Los conjuntos de la forma $\{x\} \times Y$, con $x \in X$, e $X \times \{y\}$ con $y \in Y$ se llaman *fibras* del producto $X \times Y$.

Para cada $y \in Y$, hay un mapa *inclusión* $i_y : X \rightarrow X \times Y$, dado por $i_y(x) = (x, y)$. (Análogamente, para cada $x \in X$ podemos tomar la inclusión $i_x : Y \rightarrow X \times Y$, dada por $i_x(y) = (x, y)$.) Observar que el conjunto imagen de i_y es la fibra $X \times \{y\}$.

◇ **A.15.** Las inclusiones son continuas. Concluir que las fibras son homeomorfas al factor correspondiente.

A.6. Espacios métricos

Comenzaremos estas notas recordando la definición de métrica en un espacio.

Definición A.16 (Espacio métrico). Una *métrica* en un conjunto X es una función $d : X \times X \rightarrow \mathbb{R}$ tal que

- (i) $d(x, y) \geq 0$ para todo $x, y \in X$, con $d(x, x) = 0$ y $d(x, y) > 0$ si $x \neq y$.

- (ii) $d(x, y) = d(y, x)$ para todo x, y en X .
- (iii) $d(x, y) \leq d(x, z) + d(z, y)$ para todo x, y, z en X .¹

Veamos algunos ejemplos que usaremos en estas notas.

Ejemplos A.17. 1. *Métricas en \mathbb{R}^n*

En \mathcal{R}^n podemos definir las siguientes métricas

- $d(x, y) := (\sum_{i=1}^n |x_i - y_i|^2)^{1/2}$;
- $d_1(x, y) := \sum_{i=1}^n |x_i - y_i|$;
- $d_\infty(x, y) := \max_{i=1, \dots, n} |x_i - y_i|$.

2. *Métrica inducida*

Sea (X, d) es un espacio métrico, y $A \subset X$. La restricción de d a $A \times A$ es una métrica, y por lo tanto $(A, d_{A \times A})$ es un espacio métrico.

3. *Distancia uniforme*

Sea X un conjunto, y sea $\mathcal{B}(X, \mathbb{R})$ el conjunto de funciones $f: X \rightarrow \mathbb{R}$ acotadas.² En $\mathcal{B}(X, \mathbb{R})$ definimos la métrica

$$d(f, g) := \sup_{x \in X} |f(x) - g(x)|.$$

4. *Espacios normados*

Sea V un espacio vectorial real o complejo. Una *norma* en V es una función real $\|\cdot\|: V \rightarrow \mathbb{R}$ que verifica

- (i) $\|x\| \geq 0$, y $\|x\| = 0$ si y sólo si $x = 0$ en V ;
- (ii) $\|\lambda x\| = |\lambda| \|x\|$, para todo λ escalar, y $x \in V$;
- (iii) $\|x + y\| \leq \|x\| + \|y\|$, para todo x, y en V .

Si $(V, \|\cdot\|)$ es un espacio normado entonces $d(x, y) := \|x - y\|$ es una métrica en V .

Observar que las métricas definidas en \mathbb{R}^n en el Ejemplo 1 son métricas que provienen de las normas respectivas

¹A esta propiedad se le llama *desigualdad triangular*.

²Decimos que $f: X \rightarrow \mathbb{R}$ es acotada si existe $K > 0$ tal que $|f(x)| \leq K$ para todo $x \in X$.

- $\|x\| := (\sum_{i=1}^n |x_i|^2)^{1/2}$;
- $\|x\|_1 := \sum_{i=1}^n |x_i|$;
- $\|x\|_\infty := \max_{i:1,\dots,n} |x_i|$.

A.6.1. Topología métrica

Definición A.18 (Bola). Si (X, d) es un espacio métrico, definimos la *bola de centro x y radio r* como el conjunto $B(x, r) := \{y \in X : d(x, y) < r\}$.

Definición A.19 (Topología métrica). Recordar que el conjunto de bolas de un espacio métrico forman una base de X . A la topología generada por esta base se le llama *topología métrica*.³

A.7. Completitud

La noción de sucesión convergente no es una noción intrínseca de la sucesión dado que asume la existencia de un punto particular, el límite de la sucesión. La idea es construir una noción de convergencia de sucesiones que sea intrínseca.

Definición A.20 (Sucesión de Cauchy). Una sucesión $\{x_n\}$ en un espacio métrico es una *sucesión de Cauchy* si para todo $\varepsilon > 0$ existe un N tal que $d(x_n, x_m) < \varepsilon$ para todos n y m mayores que N .

Observar que intuitivamente los términos de una sucesión de Cauchy se aproximan entre sí, a medida que n crece. (Comparar con la definición de sucesiones convergentes, donde los términos se aproximan a un mismo punto.)

A.8. Compacidad

Definición A.21 (Cubrimiento). Un *cubrimiento* de un espacio X es una familia de subconjuntos $\mathcal{U} \subset \mathcal{P}(X)$ tal que $\bigcup_{U \in \mathcal{U}} U = X$. En tal caso decimos que \mathcal{U} *cubre* a X .

³ Cuando no especifiquemos la topología en un espacio métrico significa que estamos trabajando con la topología métrica asociada.

Diremos que \mathcal{U} es un *cubrimiento abierto* si está formado por conjuntos abiertos. Un subconjunto $\mathcal{V} \subset \mathcal{U}$ que cubre a X es un *subcubrimiento* de \mathcal{U} .

Definición A.22 (Compacidad). Decimos que X es *compacto* si todo cubrimiento abierto de X tiene un subcubrimiento finito.

◇ **A.23.** Un subconjunto compacto en un espacio Hausdorff es cerrado.

Observación A.24. El resultado anterior nos permite concluir que un espacio métrico M compacto es necesariamente completo. Esto resulta de que M es un compacto es su completación \hat{M} , que es completo. Por lo tanto M es un subespacio cerrado en un completo, y por lo tanto completo.

◇ **A.25.** La intersección arbitraria de compactos, en un espacio Hausdorff, es compacta.

◇ **A.26.** Sea X un espacio compacto e Y Hausdorff. Una función $f: X \rightarrow Y$ continua es cerrada. Concluir que si f es continua y biyectiva entonces es un homeomorfismo.

Definición A.27 (Propiedad de Bolzano-Weierstrass). Decimos que un espacio X satisface la *propiedad de Bolzano-Weierstrass* si todo subconjunto infinito de X tiene un punto de acumulación en X .

◇ **A.28.** La compacidad implica la propiedad de Bolzano-Weierstrass.

Definición A.29 (Secuencialmente compacto). Un espacio X se dice *secuencialmente compacto* si toda sucesión tiene una subsucesión convergente.

◇ **A.30.** Un espacio secuencialmente compacto cumple la propiedad de Bolzano-Weierstrass.

Teorema A.31. Sea X un espacio métrico. Entonces son equivalentes:

- (i) X es compacto;
- (ii) X es satisface la propiedad de Bolzano-Weierstrass;
- (iii) X es secuencialmente compacto;

A.9. Topología cociente

Consideremos un espacio topológico (X, τ) , \sim una relación de equivalencia en X y $\pi : X \rightarrow X/\sim$ la proyección al cociente.

◇ **A.32.** La familia $\tilde{\tau} = \{A \in X/\sim : \pi^{-1}(A) \in \tau\}$ es una topología y es la más fina que hace a la proyección π continua.

A la topología del ejercicio anterior se le llama *topología cociente* de X/\sim . Observar que los abiertos en el cociente están definidos por el conjunto de clases de equivalencia tales que su unión es abierta en X .

En general si X es un espacio topológico y $f : X \rightarrow Y$ es una función podemos definir del mismo modo una topología en Y que resulta ser la más fina que hace a f continua. Esta se denomina *topología final* de Y con respecto a f .

◇ **A.33.** $\tilde{f} : X/\sim \rightarrow Y$ es continua si y sólo si $\tilde{f} \circ \pi : X \rightarrow Y$ lo es.

Propiedad universal del cociente y ejemplos

◇ **A.34. Propiedad Universal del Cociente.** Sean X e Y dos espacios topológicos, \sim una relación en X , y $f : X \rightarrow Y$ continua constante en las clases de equivalencia (i.e. $x \sim y$ entonces $f(x) = f(y)$). Entonces existe una única función continua $\tilde{f} : X/\sim \rightarrow Y$ tal que $f = \tilde{f} \circ \pi$.

◇ **A.35.** Sea X compacto e Y Hausdorff, y sea $f : X \rightarrow Y$ continua y sobreyectiva. Definimos la relación de equivalencia \sim en X declarando $x \sim y$ si y sólo si $f(x) = f(y)$. Entonces X/\sim es homeomorfo a Y .

◇ **A.36.** Sea \sim una relación de equivalencia en X tal que la proyección es abierta. Entonces X/\sim es Hausdorff si y sólo si $R = \{(x, y) \in X \times X : x \sim y\}$ es cerrado en $X \times X$.

◇ **A.37.** Se considera en \mathbb{R} la siguiente relación: $x \sim y$ si y sólo si $x - y \in \mathbb{Z}$. Entonces \mathbb{R}/\sim es homeomorfo a $[0, 1]$ identificando 0 y 1 (Sugerencia: probar que \mathbb{R}/\sim es Hausdorff).

Concluir que el cociente \mathbb{R}/\sim es homeomorfo al círculo S^1 .

◇ **A.38.** El toro n -dimensional \mathbb{T}^n se define como el cociente \mathbb{R}^n / \sim donde $(x_1, \dots, x_n) \sim (y_1, \dots, y_n)$ si y sólo si $x_i - y_i \in \mathbb{Z}$ para $i = 1, \dots, n$. \mathbb{T}^n es homeomorfo a $(S^1)^n$.

◇ **A.39.** Sea X un espacio topológico, y sea \sim una relación de equivalencia en X . Las propiedades en X de ser compacto, conexo o arcoconexo son propiedades que pasan al cociente X / \sim .

◇ **A.40.** Para $n \geq 1$ definimos $\mathbb{P}^n = S^n / \sim$, donde $x \sim y$ si y sólo si $x = y$ o $x = -y$ (i.e. se identifican puntos antipodales). El espacio \mathbb{P}^n se llama *espacio proyectivo real* de dimensión n . Observar que \mathbb{P}^n puede ser pensado como el conjunto de rectas por el origen. Probar los siguientes resultados

1. \mathbb{P}^n es compacto y Hausdorff.
2. La proyección $\pi : S^n \rightarrow \mathbb{P}^n$ es un homeomorfismo local, esto es, cada punto $x \in S^n$ tiene un entorno abierto que es mapeado homeomorfamente por π en un entorno abierto de $\pi(x)$.
3. \mathbb{P}^1 es homeomorfo al círculo S^1 .
4. \mathbb{P}^n es homeomorfo al cociente de la bola unidad cerrada $B[0, 1] \subset \mathbb{R}^n$ identificando puntos antipodales de S^{n-1} .

Geometría diferencial

En este capítulo daremos una breve introducción a los conceptos más utilizados en estas notas como forma de recordar nociones y establecer notaciones.

B.1. Variedad diferenciable

Grosso modo, una variedad es un espacio topológico que localmente se parece a \mathbb{R}^n . Con un poco más de precisión, una variedad viene provista de homeomorfismos que localmente lo identifican con abiertos de \mathbb{R}^n , y que además satisfacen cierta propiedad de compatibilidad.

Sea M un espacio topológico.

Definición B.1 (Carta local). Una *carta local* es un par (U, φ) , donde U es un abierto de M y $\varphi : U \rightarrow \mathbb{R}^n$ homeomorfismo sobre su imagen. Las componentes de $\varphi = (x_1, \dots, x_n)$ se llaman *coordenadas locales*.

El abierto $\varphi(U) \subset \mathbb{R}^n$ se le llama *entorno coordinado*: dado $p \in U \subset M$, le asignamos coordenadas $x_1(p), \dots, x_n(p)$, donde cada coordenada $x_i(p)$ es una función a valores reales definida en el entorno $U \subset M$.

Dados dos cartas locales $(U_\alpha, \varphi_\alpha)$ y (U_β, φ_β) , los *mapas de transición*, o

cambio de coordenadas son

$$\begin{aligned}\varphi_{\alpha,\beta} &:= \varphi_{\beta} \circ \varphi_{\alpha}^{-1} : \varphi_{\alpha}(U_{\alpha} \cap U_{\beta}) \rightarrow \varphi_{\beta}(U_{\alpha} \cap U_{\beta}) \\ \varphi_{\beta,\alpha} &:= \varphi_{\alpha} \circ \varphi_{\beta}^{-1} : \varphi_{\beta}(U_{\alpha} \cap U_{\beta}) \rightarrow \varphi_{\alpha}(U_{\alpha} \cap U_{\beta})\end{aligned}$$

Definición B.2 (Cartas compatibles). Dos cartas $(U_{\alpha}, \varphi_{\alpha})$ y $(U_{\beta}, \varphi_{\beta})$ son *compatibles* si sus mapas de transición son diferenciables.

Comentario B.3. Nos restringiremos al caso de funciones C^{∞} .

Definición B.4 (Atlas diferenciable). Un *atlas* para M es una colección $\mathcal{A} = \{(U_{\alpha}, \varphi_{\alpha}) : \alpha \in I\}$ de pares de cartas compatibles, tales que $M = \bigcup_{\alpha \in I} U_{\alpha}$. Diremos que un atlas es *maximal* si otro atlas que sea compatible con \mathcal{A} , este lo incluye.

Con estas definiciones previas estamos en condiciones de dar la siguiente definición.

Definición B.5 (Variedad diferenciable). Una *variedad diferenciable* de dimensión n es un espacio topológico Hausdorff, separable, con un atlas.

B.1.1. Mapas diferenciables

Las cartas locales nos permiten definir funciones diferenciables. Diremos que un mapa entre variedades diferenciables es diferenciable si lo es en coordenadas locales.

Definición B.6 (Función diferenciable). Sean M y N dos variedades diferenciables, y $F : M \rightarrow N$ una función. Diremos que F es *diferenciable* si para cualesquiera cartas (U, φ) de M y (V, ψ) de N se tiene

$$\psi \circ F \circ \varphi^{-1} : \varphi(U \cap F^{-1}(V)) \rightarrow \psi(V),$$

es diferenciable como función entre espacios euclidianos. Decimos que f es un *difeomorfismo* si F y F^{-1} son diferenciables.

B.1.2. Vectores tangentes y espacio tangente

En esta sección definiremos vectores tangentes y campo de vectores.

Existen varias formas de definir los vectores tangentes. La que tomaremos como definición es ver los vectores tangentes como una derivación.

Sea $C^\infty(M)$ el conjunto de funciones diferenciables de M en \mathbb{R} .

Definición B.7 (Vector tangente). Sea $p \in M$. Un *vector tangente* a M en p es una función $X_p : C^\infty(M) \rightarrow \mathbb{R}$ que satisface las siguientes propiedades.

1. $X_p(af + bg) = aX_p f + bX_p g$,
2. $X_p(fg) = g(p)X_p f + f(p)X_p g$,

para todo $a, b \in \mathbb{R}$ y $f, g \in C^\infty(M)$.

La propiedad 2 se llama *regla de Leibniz*.

Definición B.8 (Espacio tangente). El *espacio tangente* a una variedad M en p , al cual denotaremos por $T_p M$, es el conjunto de vectores tangentes. Este espacio está dotado de una estructura de espacio vectorial real de la siguiente manera:

$$(X_p + Y_p)(f) := X_p f + Y_p f, \quad (aX_p)(f) := aX_p f,$$

para todo $X_p, Y_p \in T_p M$ y $a \in \mathbb{R}$.

Este espacio vectorial tiene dimensión n .

Tomando coordenadas locales (U, φ) , donde x_i son las componentes de φ , resulta que el conjunto de operadores $\left. \frac{\partial}{\partial x_i} \right|_p : C^\infty(M) \rightarrow \mathbb{R}$, definido por

$$\left. \frac{\partial}{\partial x_i} \right|_p f := \frac{\partial(f \circ \varphi^{-1})}{\partial x_i}(\varphi(p)), \quad f \in C^\infty(M),$$

define vectores tangentes en $T_p M$. (El segundo miembro indica la derivada parcial de $f \circ \varphi^{-1}$, definida en un abierto de \mathbb{R}^n , respecto a la dirección i -ésima.

En particular,

$$\left\{ \left. \frac{\partial}{\partial x_i} \right|_p \in T_p M \right\}_{i=1, \dots, n}, \quad (\text{B.1})$$

definen una base de T_pM .

En particular, cuando M es un abierto de \mathbb{R}^n , los vectores tangentes $\left. \frac{\partial}{\partial x_i} \right|_p$ son exactamente los operadores derivadas parciales.

Definición B.9 (Curva). Una *curva* es una función $\alpha : I \rightarrow M$ diferenciable, definida en un intervalo abierto $I \subset \mathbb{R}$. La *velocidad de la curva* es el vector tangente $\alpha'(t) \in T_{\alpha(t)}M$, tal que

$$\alpha'(t)f = \frac{d(f \circ \alpha)}{dt}(t),$$

para todo $f \in C^\infty(M)$.

Definición B.10 (Diferencial de un mapa). Sea $F : M \rightarrow N$ una función diferenciable. Para cada $p \in M$, el *diferencial* de f en p está dado por la función

$$DF(p) : T_pM \rightarrow T_{f(p)}N,$$

definido por

$$DF(p)(X_p)f = X_p(f \circ F),$$

para todo $v \in T_pM$, y $f \in C^\infty(N)$.

El siguiente resultado muestra que la base dadas por las derivadas parciales canónica de \mathbb{R}^n se corresponden con la base (B.1) en T_pM por medio de una carta local.

Lema B.11. Sea $p \in M$ y $(U, \varphi = (x_i))$ un carta local. Entonces,

$$D\varphi(p) \left. \frac{\partial}{\partial x_i} \right|_p = \left. \frac{\partial}{\partial x_i} \right|_{\varphi(p)},$$

donde $\left\{ \left. \frac{\partial}{\partial x_i} \right|_x \right\}$ son la base definida en (B.1), para $x \in \varphi(U)$, y donde $\left\{ \left. \frac{\partial}{\partial x_i} \right|_{\varphi(p)} \right\}$ representa la base canónica de $T_x\mathbb{R}^n$. \square

◇ **B.12.** Probar el resultado.

Comentario B.13. Sea $U \subset \mathbb{R}^n$ y $p \in U$. Una base del espacio tangente T_pU es el conjunto de (operadores) derivadas parciales. Identificando las derivadas parciales con los vectores e_i de la base canónica de \mathbb{R}^n , es fácil ver que existe una identificación (un isomorfismo) entre \mathbb{R}^n y T_pU , a saber,

$$v = \sum_{i=1}^n v_i e_i \in \mathbb{R}^n \mapsto \left. \frac{\partial}{\partial v} \right|_p = \sum_{i=1}^n v_i \left. \frac{\partial}{\partial x_i} \right|_p \in T_pU.$$

Luego, utilizando Lema B.11, podemos identificar $T_pM \cong \mathbb{R}^n$. Esta identificación nos permite “visualizar” los vectores tangentes a una variedad diferenciable como vectores.

A continuación mostraremos una forma de calcular derivadas a través de curvas.

Proposición B.14. *Sea $F : M \rightarrow N$ una función diferenciable y sea $p \in M$ y $X_p \in T_pM$. Si $\alpha : I \rightarrow M$ es una curva tal que $\alpha(0) = p$ y $\alpha'(0) = X_p$, entonces se tiene*

$$DF(p)X_p = \left. \frac{d}{dt}(F \circ \alpha) \right|_{t=0}.$$

Observar que todo vector $X \in T_pM$ puede ser expresado como

$$X_p = \sum_{i=1}^n X_p^{(i)} \left. \frac{\partial}{\partial x_i} \right|_p,$$

donde X_i son las componentes del vector. Además, si X_p es conocido podemos recuperar sus componentes actuando sobre las funciones coordenadas $x_i = \varphi_i(x) : U \rightarrow \mathbb{R}$. Resulta que

$$X_p x_k = \sum_{i=1}^n X_p^{(i)} \left. \frac{\partial}{\partial x_i} \right|_p x_k = \sum_{i=1}^n X_p^{(i)} \left. \frac{\partial \pi_k}{\partial x_i} \right|_p (\varphi(p)) = X_p^{(k)}$$

donde $\pi_k = x_k \circ \varphi^{-1}$ es la proyección canónica en la coordenada k -ésima.

B.1.3. Cambio de coordenadas

Con estas notaciones podemos escribir una base del tangente en función de otra mediante el diferencial del cambio de coordenadas $\phi \circ \varphi^{-1} : \varphi(U \cap V) \rightarrow \psi(U \cap V)$.

Si escribimos para simplificar la notación $y(x) = \psi \circ \varphi^{-1}(x)$, y su inversa $x(y) = \varphi \circ \psi^{-1}(y)$, entonces resulta del cálculo del jacobiano que

$$\left. \frac{\partial}{\partial x_i} \right|_p = \sum_{k=1}^n \left. \frac{\partial y_k}{\partial x_i} \right|_p (\varphi(p)) \left. \frac{\partial}{\partial y_k} \right|_p.$$

$$\left. \frac{\partial}{\partial y_i} \right|_p = \sum_{k=1}^n \left. \frac{\partial x_k}{\partial y_i} \right|_p (\psi(p)) \left. \frac{\partial}{\partial x_k} \right|_p.$$

B.2. Fibrado tangente

Definición B.15 (Fibrado tangente). Sea M una variedad diferenciable. El *fibrado tangente* de M es el conjunto

$$TM = \bigcup_{p \in M} T_p M = \{(p, v_p) : p \in M, v_p \in T_p M\}.$$

El fibrado tangente TM puede dotarse de una estructura de variedad diferenciable de la siguiente manera.

Definición B.16 (Proyección canónica). Sea $\pi_M : TM \rightarrow M$ la *proyección canónica* : $\pi_M(p, v_p) = p$, para todo $(p, v_p) \in TM$.

Observar que para cada $p \in M$, $\pi_M^{-1}(p) = T_p M$. En términos de fibrados vectoriales $\pi_M^{-1}(p)$ se llama *fibra* sobre p .

Dada una carta local $(U_\alpha, \varphi_\alpha)$ de M , consideremos el par $(\tilde{U}_\alpha, \tilde{\varphi}_\alpha)$ donde

$$\tilde{\varphi}_\alpha : \pi_M^{-1}(U_\alpha) \rightarrow \mathbb{R}^{2n}, \quad \tilde{\varphi}_\alpha \left(x, \sum_{i=1}^n v_i \frac{\partial}{\partial x_i} \Big|_x \right) = (\varphi_\alpha(x), v_1, \dots, v_n).$$

La idea es utilizar estos mapas para definir la topología y estructura diferenciable de TM . Un conjunto $A \subset TM$ es abierto si y sólo si $\tilde{\varphi}_\alpha(A \cap \pi_M^{-1}(U_\alpha))$ es abierto en \mathbb{R}^{2n} para cada α . De esta manera los mapas $\tilde{\varphi}_\alpha$ son homeomorfismos entre abiertos de TM y abiertos de \mathbb{R}^{2n} . Estos mapas definen un atlas compatible dado que los cambios de coordenadas involucran a los jacobianos de los cambios de coordenadas de cartas en M (cf. Sección B.1.3).

B.3. Campo de vectores

Definición B.17 (Campo de vectores). Un *campo de vectores* X en una variedad M es una función que a cada $p \in M$ le corresponde un vector $X_p \in T_p M$. En este sentido un campo de vectores X es una sección de TM , i.e. $X : M \rightarrow TM$ tal que $\pi_M \circ X = \text{Id}_M$.

Si X es un campo vectorial y $f \in C^\infty(M)$, entonces Xf es la función definida sobre M a valores reales tales que

$$Xf(p) = X_p f, \quad p \in M.$$

Diremos que X es un *campo vectorial diferenciable*, y escribiremos $X \in \mathcal{X}(M)$, si $Xf \in C^\infty(M)$ para todo $f \in C^\infty(M)$.

Comentario B.18. La noción de diferenciabilidad se puede dar en coordenadas locales. Esto significa que si escribimos X en coordenadas locales, entonces los coeficientes del campo son diferenciables. Precisamente, si $X_p = \sum_{i=1}^n \alpha_i \frac{\partial}{\partial x_i} \Big|_p$, cuando p se mueve en un entorno U de M , los coeficientes α_i están bien definidos como funciones de U en \mathbb{R} . Resulta que X es diferenciable si los coeficientes α_i lo son. Es fácil ver que esta definición es independiente de las coordenadas locales.

Comentario B.19. Existe otra forma equivalente de dar la definición de diferenciabilidad de un campo de vectores a través del fibrado tangente. El espacio TM puede dotarse de una estructura de variedad diferenciable (ver Sección B.2). Luego un campo de vectores $X : M \rightarrow TM$ es diferenciable si lo es como función entre variedades.

Un campo $X \in \mathcal{X}(M)$ induce un mapa $X : C^\infty(M) \rightarrow C^\infty(M)$, que manda f en Xf . Este mapa cumple las propiedades de ser una *derivación*

1. $X(af + bg) = aXf + bXg$, $a, b \in \mathbb{R}$
2. $X(fg) = gXf + fXg$. (Regla de Leibniz)

Recíprocamente, toda derivación $D : C^\infty(M) \rightarrow C^\infty(M)$ se puede inducir por un campo vectorial diferenciable. Basta definir $X : M \rightarrow TM$, de manera tal que $X_p f := D(f)(p)$, para todo $f \in C^\infty(M)$.

Comentario B.20. A lo largo de estas notas usaremos indistintamente la notación X_p y $X(p)$ para indicar el valor en $T_p M$ de un campo vectorial X .

B.3.1. Corchete de Lie

Ver los campos de vectores diferenciables como derivaciones nos permite componer campos. Esto es, $XY : C^\infty(M) \rightarrow C^\infty(M)$, dado por $XY(f) = X(Y(f))$. Esta composición induce naturalmente un campo lineal y por lo tanto satisface la propiedad 1. Sin embargo no induce en general una derivación (no satisface la regla de Leibniz 2).

Definición B.21 (Corchete de Lie). Dados dos campos $X, Y \in \mathcal{X}(M)$, definimos el *corchete de Lie* como

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

◇ **B.22.** Probar que si $X, Y \in \mathcal{X}(M)$ entonces $[X, Y]$ define una derivación y por lo tanto $[X, Y] \in \mathcal{X}(M)$.

Para cada $p \in M$, el corchete $[X, Y]$ está dado por

$$[X, Y]_p f = X_p(Yf) - Y_p(Xf).$$

Se puede probar que el corchete de Lie verifica las siguientes propiedades:

1. $[X, Y] = -[Y, X]$ (antisimetría)
2. $[aX + bY, Z] = a[X, Z] + b[Y, Z]$, $a, b \in \mathbb{R}$ (bilineal)
3. $[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0$, (identidad de Jacobi)

Definición B.23 (Álgebra de Lie). Un espacio vectorial V con una operación $[\cdot, \cdot] : V \times V \rightarrow V$ que satisface las propiedades (i), (ii) y (iii) se denomina *álgebra de Lie*.

Con estas definiciones podemos concluir que el conjunto $\mathcal{X}(M)$, dotado del corchete de Lie, es un álgebra de Lie.

B.3.2. Ecuaciones diferenciales y subgrupos a un parámetro

Sea $X : M \rightarrow TM$ un campo de vectores diferenciable. Una *curva integral* de X es una curva $\alpha : I \rightarrow M$, definida en un intervalo $I \subset \mathbb{R}$, tal que $\alpha(t)$ es una solución de

la ecuación diferencial en M

$$\dot{\alpha}(t) = X(\alpha(t)), \quad t \in I. \quad (\text{B.2})$$

En coordenadas locales, la ecuación (B.2) es una ecuación diferencial ordinaria en un abierto de \mathbb{R}^n (siendo n la dimensión de M). Entonces, por el teorema de existencia y unicidad de ecuaciones diferenciales ordinarias, resulta que existe una única curva integral solución de la ecuación (B.2). En el siguiente resultado resumimos este resultado.

Teorema B.24. *Sea $X \in \mathcal{X}(M)$ un campo campo diferenciable, y $p \in M$. Entonces existe un abierto U de p , un intervalo I que contiene a 0 , y un mapa diferenciable $\phi : I \times U \rightarrow M$ tal que $\alpha_q : I \rightarrow M$ dada por $\alpha_q(t) = \phi(t, q)$, (con $q \in U$) es la única curva que satisface*

$$\dot{\alpha}_q(t) = X(\alpha_q(t)), \quad \alpha_q(0) = q.$$

□

Fijado el parámetro t , la función $\phi_t : U \rightarrow M$ dada por $\phi_t(x) = \phi(t, x)$, se llama *flujo local* el cual cumple las siguientes propiedades:

1. ϕ_0 es la identidad de U ,
2. $\phi_s \circ \phi_t = \phi_{s+t}$, para todo $s, t \in U$,
3. cada ϕ_t es un difeomorfismo con $(\phi_t)^{-1} = \phi_{-t}$

Cuando ϕ_t está definida para todo $t \in \mathbb{R}$ decimos que el campo de vectores $X \in \mathcal{X}(M)$ es *completo*. En particular, si M es compacto entonces cualquier $X \in \mathcal{X}(M)$ es completo.

Definición B.25 (Grupo a un parámetro). Si $X \in \mathcal{X}(M)$ es completo, entonces el flujo asociado $\phi_t, t \in \mathbb{R}$, es un *grupo a un parámetro* de difeomorfismos. Esto es

$$\phi_0 = \text{Id}_M, \quad \phi_s \circ \phi_t = \phi_{s+t}, \quad (\phi_t)^{-1} = \phi_{-t},$$

para todo $s, t \in \mathbb{R}$.

B.4. Subvariedades

Una subvariedad de una variedad diferenciable es un subconjunto de M que poseé la estructura diferenciable de M .

Definición B.26 (Subvariedad). Sea M una variedad diferenciable de dimensión n . Un subconjunto $S \subset M$ es una *subvariedad* si existe una carta local (U, φ) de M , en un entorno de $p \in S$, tal que

$$\varphi : U \rightarrow \mathbb{R}^k \times \mathbb{R}^{n-k}, \quad \varphi(S \cap U) = \varphi(U) \cap (\mathbb{R}^k \times \{0\})$$

para algún $0 \leq k \leq n$. A un par (U, φ) como antes lo llamamos *cartas local con la propiedad de subvariedad*.

En particular, S es una variedad diferenciable en sí misma (con la topología inducida como subespacio de M), tomando como atlas las cartas locales $(S \cap U, \varphi|_{S \cap U})$ donde (U, φ) es carta local de M con la propiedad de subvariedad.

El espacio tangente $T_p S$ de la subvariedad S en $p \in S$, es considerado como subespacio vectorial de $T_p M$.

Definición B.27 (Valor regular). Sea $f : M \rightarrow N$ una función diferenciable entre variedades. Decimos que q es un *valor regular* si $Df(p)$ es sobreyectivo para todo $p \in f^{-1}(q)$. En otro caso diremos que $q \in N$ es un *valor crítico*.

El siguiente resultado es una herramienta excelente para mostrar que ciertos conjuntos son (sub)variedades.

Teorema B.28 (Teorema de la submersión). Sea $f : M \rightarrow N$ una función diferenciable entre variedades. Supongamos que q es un valor regular de f . Entonces $S := f^{-1}(q)$ es una subvariedad de M de dimensión $\dim M - \dim N$, y el tangente está dado por

$$T_p S = \ker Df(p), \quad \forall p \in S.$$

□

Comentario B.29. Pueden existir subconjuntos de M que sean variedades diferenciables pero que no sean subvariedades. En particular esto ocurre cuando el subconjunto tiene una topología diferente a la inducida por M .

Definición B.30 (Inmersión, submersión, encaje). Sea $f : M \rightarrow N$ una función diferenciable.

1. f es una *inmersión* si $Df(x) : T_x M \rightarrow T_{f(x)} N$ es inyectiva en todo $x \in M$.
2. f es una *submersión* si $Df(x) : T_x M \rightarrow T_{f(x)} N$ es sobreyectiva en todo $x \in M$.
3. f es una *encaje* si f es una inmersión, inyectiva, de manera tal que f induce un homeomorfismo entre M y $f(M)$, donde $f(M)$ tiene la topología como subespacio de N .

Definición B.31 (Subvariedad inmersa). Si $f : M \rightarrow N$ es una inmersión inyectiva, entonces $f(M)$ es una *subvariedad inmersa*.

A una subvariedad inmersa se le puede dar una estructura de variedad diferenciable. Una forma de introducir una estructura diferenciable es via el mapa $f : M \rightarrow f(M)$. Esto es, definimos una topología en $f(M)$ de manera tal que $f : M \rightarrow f(M)$ sea un homeomorfismo, para luego usar la estructura diferenciable de M para inducir una estructura diferenciable en $f(M)$.

Comentario B.32. Es importante destacar que una subvariedad inmersa $f(M)$ puede no ser una subvariedad de N . La razón es que la topología de la subvariedad inmersa puede no coincidir con la topología inducida como subvariedad. Un ejemplo de esto es la “rotación irracional en el toro” $\mathbb{T}^2 = S^1 \times S^1$. (Ver Ejemplo C.5.)

B.5. Puntos críticos

Sea M una variedad diferenciable, y $f : M \rightarrow \mathbb{R}$ una función $C^2 = C^2(M)$. Diremos que una función tiene un *punto crítico* en $p \in M$ si la derivada $Df(p) : T_p M \rightarrow \mathbb{R}$ es la aplicación nula (donde $T_p M$ denota el espacio tangente de M en p). Denotaremos por $\mathcal{C}(f)$ al conjunto de puntos críticos de f .

Una subvariedad $N \subset M$ de dimensión l conexa, se dice *subvariedad crítica* si $N \subset \mathcal{C}(f)$.

B.6. Hessiano

Si $p \in M$ es un punto crítico de una aplicación $f : M \rightarrow \mathbb{R}$, $f \in C^2$, entonces se define el *Hessiano* de f en p como la forma bilineal simétrica

$$H_p(f) : T_p M \times T_p M \rightarrow \mathbb{R},$$

dada en la base coordenada por (U, φ) ,

$$H_p(f) \left(\left. \frac{\partial}{\partial x_i} \right|_p, \left. \frac{\partial}{\partial x_j} \right|_p \right) = \frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j} (\varphi(p)), \quad i, j = 1, \dots, n,$$

siendo $\tilde{f} = f \circ \varphi^{-1}$.

Lema B.33. *La definición del Hessiano en un punto crítico no depende de las coordenadas locales.*

Demostración. Sea (ψ, V) otra carta local tal que $p \in V$. Observar que usando el cambio de coordenadas dado más arriba resulta que

$$H_p(f) \left(\left. \frac{\partial}{\partial y_i} \right|_p, \left. \frac{\partial}{\partial y_j} \right|_p \right) = \sum_{k, k'=1}^n \frac{\partial x_k}{\partial y_i} (\psi(p)) \frac{\partial x_{k'}}{\partial y_j} (\psi(p)) \frac{\partial^2 f \circ \varphi^{-1}}{\partial x_i \partial x_j} (\varphi(p)).$$

Luego, para que la definición sea consistente se tiene que probar que

$$\frac{\partial^2 f \circ \psi^{-1}}{\partial y_i \partial y_j} (\psi(p)) = \sum_{k, k'=1}^n \frac{\partial x_k}{\partial y_i} (\psi(p)) \frac{\partial x_{k'}}{\partial y_j} (\psi(p)) \frac{\partial^2 f \circ \varphi^{-1}}{\partial x_i \partial x_j} (\varphi(p)).$$

Usando la regla de la cadena resulta, si $s : \psi(V) \rightarrow \mathbb{R}$, es diferenciable, y escribimos x como función de y , entonces

$$\frac{\partial s}{\partial y_i} (y) = \sum_{k=1}^n \frac{\partial s}{\partial x_k} (x) \frac{\partial x_k}{\partial y_i} (y) \quad (\text{B.3})$$

Luego, resulta de (B.3), para $s = f \circ \psi^{-1}$,

$$\begin{aligned} \frac{\partial^2 f \circ \psi^{-1}}{\partial y_j \partial y_i} (y) &= \sum_{k=1}^n \frac{\partial}{\partial y_j} \left[\frac{\partial f \circ \psi^{-1}}{\partial x_k} (x) \frac{\partial x_k}{\partial y_i} (y) \right] \\ &= \sum_{k=1}^n \frac{\partial}{\partial y_j} \left(\frac{\partial f \circ \psi^{-1}}{\partial x_k} (x) \right) \frac{\partial x_k}{\partial y_i} (y) + \frac{\partial f \circ \psi^{-1}}{\partial x_k} (x) \frac{\partial}{\partial y_j} \frac{\partial x_k}{\partial y_i} (y). \end{aligned}$$

Utilizando nuevamente (B.3) tenemos

$$\frac{\partial}{\partial y_j} \left(\frac{\partial f \circ \psi^{-1}}{\partial x_k}(x) \right) = \sum_{k'=1}^n \frac{\partial^2 f \circ \psi^{-1}}{\partial x_{k'} \partial x_k}(x) \frac{\partial x_{k'}}{\partial y_j}(y),$$

de donde concluimos

$$\frac{\partial^2 f \circ \psi^{-1}}{\partial y_j \partial y_i}(y) = \sum_{k,k'=1}^n \frac{\partial^2 f \circ \psi^{-1}}{\partial x_{k'} \partial x_k}(x) \frac{\partial x_{k'}}{\partial y_j}(y) \frac{\partial x_k}{\partial y_i}(y) + \frac{\partial f \circ \psi^{-1}}{\partial x_k}(x) \frac{\partial}{\partial y_j} \frac{\partial x_k}{\partial y_i}(y).$$

El resultado sigue de observar que el último miembro es cero en el punto crítico. \square

Comentario B.34. Si p no es un punto crítico de f , entonces el Hessiano no está bien definido dado que dependería de la carta empleada. Hay formas de extender el hessiano a puntos no críticos pero se requiere dar una estructura de métrica Riemanniana.

Comentario B.35. Hay otra forma de probar lo anterior usando el corchete de Lie. Sean X y Y dos campos de vectores C^∞ . Observar que Xf define un mapa C^∞ en M : a cada $p \in M$, $Xf(p)$ da la derivada direccional de f en la dirección del vector $X(p)$. Luego tiene sentido tomar $Y(Xf)$.

Afirmación 1: Si p es un punto crítico de f entonces $X(p)(Yf) = Y(p)(Xf)$. Esto resulta de que el corchete de Lie define un nuevo campo de vectores, y entonces

$$[X, Y](p)f = Df(p)[X, Y](p) = 0.$$

Afirmación 2: Si X' y Y' son campos que coinciden en p con X y Y respectivamente, entonces $X'(p)(Y'f) = X'(p)(Y'f)$.

◇ **B.36.** Probar la afirmación.

Luego, se puede definir el Hessiano de f en p como $H_p(f)(X(p), Y(p)) = X(p)(Y'f)$. Observar que con esta definición $H_p(f)$ es un mapa bilinear simétrico.

◇ **B.37.** Mostrar que con esta definición resulta $H_p(f) \left(\frac{\partial}{\partial x_i} \Big|_p, \frac{\partial}{\partial x_j} \Big|_p \right) = \frac{\partial^2 \tilde{f}}{\partial x_i \partial x_j}$, siendo $\tilde{f} = f \circ \varphi^{-1}$.

Dada una subvariedad crítica $N \subset M$, consideramos un fibrado vectorial normal η en N . Esto es $\eta(x)$ es un subespacio complementario a $T_x N$ en $T_x M$, i.e.

$$T_x M \simeq T_x N \oplus \eta(x).$$

En el caso de haber una métrica Riemanniana, podemos tomar $\eta(x) = T_x N^\perp \subset T_x M$. Observar que el Hessiano de f en p , tiene la propiedad de que toma el valor cero si nos restringimos a $T_x N$. Luego $H_p(f)$ determina una forma bilineal simétrica en el fibrado normal

$$H_p^\perp : \eta(p) \times \eta(p) \rightarrow \mathbb{R}.$$

Decimos que $N \subset M$ es una subvariedad crítica no degenerada si $H_p^\perp(f)$ es no singular (como forma cuadrática).

B.7. Funciones de Morse, Morse-Bott

Definición B.38 (Función de Morse-Bott). Sea $f : M \rightarrow \mathbb{R}$ diferenciable, se dice que f es de *Morse-Bott* si $\mathcal{C}(f) = \bigcup_j N_j$ con N_j subariedades críticas no degeneradas disjuntas. En particular si N_j son puntos, decimos que f es de Morse.

Como $H_p^\perp(f)$ es no degenerada, sus valores propios (reales) son no nulos, por lo tanto podemos descomponer $\eta(p)$ en dos subespacios $\eta(p)^+$ y $\eta(p)^-$ tales que $\eta(p) = \eta(p)^+ \oplus \eta(p)^-$ y $H_p^\perp(f)$ es positivo sobre $\eta(p)^+$ y negativo sobre $\eta(p)^-$.

Definición B.39 (Índice de subvariedad crítica). Sea $N \subset M$ una subvariedad crítica, se define el *índice* de N como la dimensión de $\eta(p)^-$ con $p \in N$.

Observación B.40. Observar que por continuidad la dimensión de $\eta(p)^-$ tiene que ser constante $\forall p \in N$ dado que $H_p^\perp(f)$ es no degenerado.

Lema B.41 (de Morse). Sean $f : M \rightarrow \mathbb{R}$ de clase C^2 y N subvariedad crítica no degenerada de índice λ . Dado $a \in N$ existe (φ, U) carta local tal que

$$a) \quad \varphi : U \cap N \rightarrow \mathbb{R}^l \times \{0\}^{n-l}$$

$$b) \quad f \circ \varphi^{-1}(x_0, x_1, x_2) = f(a) + \|x_2\|^2 - \|x_1\|^2, \text{ con } x_0 \in \mathbb{R}^l, x_1 \in \mathbb{R}^\lambda, x_2 \in \mathbb{R}^{n-\lambda-l}.$$

□

B.8. Métrica Riemanniana

Definición B.42 (Métrica Riemanniana). Sea M una variedad diferenciable, y sea TM el fibrado tangente. Una *métrica Riemanniana* es una familia de productos internos $\langle \cdot, \cdot \rangle_x$, $x \in M$, definidos en cada espacio tangente $T_x M$, de manera que tal que la dependencia en x es diferenciable.

En otras palabras, una métrica Riemanniana en M es una función que a cada $x \in M$ le corresponde una forma bilineal simétrica definida positiva en $T_x M$.

La condición de diferenciabilidad de los productos internos se puede establecer de la siguiente manera. El mapa $x \in M \mapsto \langle \cdot, \cdot \rangle_x$ es diferenciable, si para cualesquiera campos $X, Y \in \mathcal{X}(M)$, se tiene

$$x \in M \mapsto \langle X_x, Y_x \rangle_x \in \mathbb{R},$$

es una función diferenciable.

Definición B.43 (Variedad Riemanniana). Una *variedad Riemanniana* es una variedad diferenciable dotado de una métrica Riemanniana.

Es un resultado clásico de geometría Riemanniana que existen métricas Riemannianas sobre cualquier variedad diferenciable.

Comentario B.44. Utilizaremos indistintamente la notación (M, g) o $(M, \langle \cdot, \cdot \rangle_x)$ para indicar que M es una variedad Riemanniana.

Ejemplo B.45. En \mathbb{R}^n , dar una métrica Riemanniana es equivalente a dar una familia de matrices simétricas definidas positivas $Q: \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$. (Esto resulta de que dar un producto interno es equivalente a dar una matriz definida positiva.)

B.8.1. Métrica inducida por inmersión

Si $f: M \rightarrow N$ es una inmersión, entonces dada una métrica Riemanniana en N induce una métrica Riemanniana en M (el pullback de la métrica). Basta definir

$$\langle v, w \rangle_x := \langle Df(x)v, Df(x)w \rangle_{f(x)}, \quad x \in M,$$

y para todo $v, w \in T_x M$.

En particular, si $M \subset N$ es una subvariedad, y f es la inclusión, entonces cualquier métrica Riemanniana en N induce una métrica Riemanniana en M . De la definición anterior, el producto interno en TM está dado trivialmente por la restricción del producto interno de TN a TM .

B.9. Gradiente

Sea $(M, \langle \cdot, \cdot \rangle_x)$ una variedad Riemanniana. Sea $\phi : M \rightarrow \mathbb{R}$ una función $C^\infty(M)$. El diferencial de esta ϕ puede ser pensado como un campo de vectores “duales”, i.e. es una sección $D\phi : M \rightarrow TM^*$ donde

$$D\phi(x) \in T_x M^*,$$

siendo $T_x M^*$ el espacio dual de $T_x M$.

Digresión Sea V es un espacio con producto interno $\langle \cdot, \cdot \rangle$. Si $\varphi \in V^*$, i.e. $\varphi : V \rightarrow \mathbb{R}$ es una funcional lineal, entonces por el teorema Riesz se tiene que existe $v_\varphi \in V$ tal que

$$\varphi(w) = \langle v_\varphi, w \rangle, \quad \forall w \in V$$

La digresión anterior motiva la siguiente definición.

Definición B.46 (Gradiente). Sea $(M, \langle \cdot, \cdot \rangle_x)$ una variedad Riemanniana, y $\phi : M \rightarrow \mathbb{R}$ una función diferenciable. Definimos el *gradiente* de ϕ como el vector tangente $\text{grad } \phi(x) \in T_x M$, $x \in M$, que satisface

$$D\phi(x)v_x = \langle \text{grad } \phi(x), v_x \rangle, \quad \forall v_x \in T_x M.$$

El gradiente induce el *campo vectorial gradiente* definido por $\text{grad } \phi : M \rightarrow TM$

Comentario B.47. Es importante destacar que el gradiente de ϕ depende de la métrica Riemanniana empleada.

Lema B.48. Si $\phi : \mathbb{R}^n \setminus \{0\} \rightarrow \mathbb{R}$ es diferenciable y $\text{grad } \phi(x)$ es el gradiente para la métrica $\langle \zeta, \eta \rangle_x = \eta^T Q(x) \zeta$, (cf. Ejemplo B.45), se tiene

$$\text{grad } \phi(x) = Q(x)^{-1} \nabla \phi(x).$$

Demostración. Se tiene por un lado

$$D\phi(x)\dot{x} = \langle \text{grad } \phi(x), \dot{x} \rangle_x = \dot{x}^T Q(x) \text{grad } \phi(x)$$

a su vez

$$D\phi(x)\dot{x} = \langle \nabla \phi(x), \dot{x} \rangle = \dot{x}^T \nabla \phi(x) = \dot{x}^T Q(x) Q(x)^{-1} \nabla \phi(x)$$

de donde resulta

$$\text{grad } \phi(x) = Q(x)^{-1} \nabla \phi(x).$$

□

Lema B.49. Sea $\phi : N \rightarrow \mathbb{R}$ diferenciable en M variedad Riemanniana, y sea $M \subset N$ una subvariedad con la métrica Riemanniana inducida (ver Sección B.8.1). Si $x \in M$, entonces el $\text{grad}|_M(x)$ de la restricción $\phi_M : M \rightarrow \mathbb{R}$ es la proyección ortogonal de $\text{grad } \phi(x) \in T_x M$ en $T_x M$. □

◇ **B.50.** Dar una prueba de este lema.

Corolario B.51. Sea $M \subset \mathbb{R}^n$ una subvariedad diferenciable, con la estructura Riemanniana inducida por la métrica Euclídea en \mathbb{R}^n . El gradiente de una función $\phi : M \rightarrow \mathbb{R}$, que tiene una extensión $\tilde{\phi}$ a un abierto Euclídeo de M , puede obtenerse como la proyección ortogonal del gradiente de la extensión en \mathbb{R}^n en el espacio tangente $T_x M$. Esto es, si denotamos por $\nabla \tilde{\phi}(x)$ al gradiente de $\tilde{\phi}$ en x (con métrica Euclídea), entonces

$$\text{grad}|_M \phi(x) = \pi_{T_x M} \nabla \tilde{\phi}(x),$$

siendo $\pi_{T_x M} : \mathbb{R}^n \rightarrow T_x M$ la proyección ortogonal.

B.10. Flujo gradiente

Sean (M, g) una variedad Riemanniana y $\phi : M \rightarrow \mathbb{R}$ una función diferenciable, tenemos entonces $\text{grad } \phi : M \rightarrow TM$ y se define la ecuación diferencial $\dot{x}(t) = -\text{grad } \phi(x(t))$. Si $x(t)$ es una solución, resulta

$$\frac{d}{dt} \phi(x(t)) = D\phi(x(t))\dot{x}(t) = \langle \text{grad } \phi(x(t)), \dot{x}(t) \rangle = -\|\text{grad } \phi(x(t))\|^2 < 0$$

siempre que $x(t)$ no sea punto crítico de ϕ . Es decir, ϕ es decreciente en las líneas del flujo.

Teorema B.52. Sean M una variedad Riemanniana compacta y $f : M \rightarrow \mathbb{R}$ una función C^2 . Entonces la solución $\gamma_x(t)$ de $\dot{\gamma} = -\text{grad} f(\gamma)$, con condición inicial $\gamma_x(0) = x$ están definidas $\forall t \in \mathbb{R}$ y además $\lim_{t \rightarrow +\infty} \gamma_x(t)$ y $\lim_{t \rightarrow -\infty} \gamma_x(t)$ son puntos críticos de f .

Demostración. Sea $\gamma_x : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}$ solución de $\dot{\gamma} = -\text{grad} f(\gamma)$ con $\gamma_x(0) = x$, al ser M compacta, podemos extender las soluciones de modo que estén definidas para todo $t \in \mathbb{R}$.

Sea γ una solución genérica, resulta que

$$f \circ \gamma(b) - f \circ \gamma(a) = \int_a^b \frac{d}{dt}(f \circ \gamma)(t) dt.$$

Como M es compacta tenemos que $\text{Im}(f \circ \gamma)$ es acotada, luego se tiene que $|\frac{d}{dt} f \circ \gamma(t)|$ está acotado y $\lim_{t \rightarrow \pm\infty} |\frac{d}{dt} f \circ \gamma(t)| = 0$. En particular,

$$\lim_{t \rightarrow \pm\infty} \|\text{grad} f(\gamma(t))\| = 0.$$

Sea U una unión de entornos abiertos disjuntos de los puntos críticos. Luego $M \setminus U$ es compacto y por lo tanto $\|\text{grad} f(x)\| \geq \alpha_0 > 0$, $\forall x \in M \setminus U$, es decir tiene mínimo. Luego usando que $\|\text{grad} f(\gamma(t))\|$ tiende a cero, para t_0 suficientemente grande, $\gamma(t) \in U_j \forall t \geq t_0$, siendo U_j una componente conexa de U . Si ahora nos tomamos una sucesión decreciente tales U , obtenemos que $\gamma(t)$ tiende a un punto crítico de f . \square

Comentario B.53. El resultado anterior se puede extender para el caso M no compacta pero pidiéndole a f que tenga subniveles compactos, i.e. $f^{-1}(-\infty, a]$ sea compacta para todo $a \in \mathbb{R}$. Observar que en tal caso, $\gamma(t)$ existe para todo $t \in \mathbb{R}^+$. Además $\gamma(t)$ converge a una componente conexa del conjunto de puntos críticos. Es decir, se sigue cumpliendo el teorema anterior solo que obtengo información a futuro.

Comentario B.54. Si M es compacta y $f : M \rightarrow \mathbb{R}$ es de Morse-Bott, entonces $\gamma(t)$ converge a un único punto crítico. En efecto, del teorema anterior sabemos que $\gamma(t)$ “converge” a una componente conexa de $\mathcal{C}(f)$, i.e. $\gamma_x(t) \rightarrow N_j$ para algún j . Luego, el ω -límite de x esta contenido en N_j . Sea $p \in \omega$ -límite. Tomando coordenadas locales de Morse-Bott, es decir, un entorno abierto $U \subset M$ de p tal que $\varphi(U \cap N_j) =$

$\mathbb{R}^l \times \{0\}^{n-l}$ y $f \circ \varphi^{-1}(x_1, x_2, x_3) = \|x_2\|^2 - \|x_3\|^2$, con $x_1 \in \mathbb{R}^l$, $x_2 \in \mathbb{R}^{n_+}$ y $x_3 \in \mathbb{R}^{n_-}$ siendo $l + n_+ + n_- = n$ (asumimos $f(p)=0$). Luego en coordenadas locales y para una métrica adaptada¹, el flujo gradiente se describe como las soluciones de

$$\begin{cases} \dot{x}_1 = 0 \\ \dot{x}_2 = -x_2 \\ \dot{x}_3 = x_3 \end{cases}$$

Luego, en el caso en que la condición inicial fuera $x_3 \neq 0$ eventualmente se escapará del abierto U y tendería a otra componente de $\mathcal{C}(f)$. Por lo tanto en estas coordenadas $\gamma_x(t)$ es de la forma $(x_1, x_2, 0)$ y por lo tanto

$$\lim_{t \rightarrow +\infty} \varphi(\gamma_x(t)) = (x_1, 0, 0).$$

Análogo para $t \rightarrow -\infty$.

Comentario B.55. Este argumento sugiere que si hay una única variedad crítica de índice 0, entonces el conjunto de condiciones iniciales x donde el flujo no converge a dicha subvariedad tiene medida cero.

¹Por una métrica adaptada nos referimos a una perturbación local de la métrica. En nuestro caso podemos hacer una homotopía, en coordenadas locales, entre la métrica usual y la Euclídea de manera tal que sólo afecte a la métrica en un entorno del punto. Si $\phi : \varphi(U) \rightarrow \mathbb{R}$ es una función chichón que toma el valor 1 en una bola coordenadas pequeña (puntos con coordenadas x_1, \dots, x_n tales que $x_1 + \dots + x_n^2 < r$) y que tome el valor 0 fuera de una bola mayor – todo dentro de $\varphi(U)$. Luego se define $g = g_0(x)(1 - \phi(x)) + g_{\text{can}}(x)\phi(x)$, en U y se extiende como g_0 fuera, siendo g_0 la métrica de M .

Grupos de Lie

Sea $\mathbb{R}^{n \times n}$ el espacio de las matrices $n \times n$ con coeficientes reales con la identificación $\mathbb{R}^{n \times n} \cong \mathbb{R}^{n^2}$, convirtiendo $\mathbb{R}^{n \times n}$ en un espacio topológico. El subconjunto $\text{Gl}(n, \mathbb{R}) \subset \mathbb{R}^{n \times n}$ de las matrices $A \in \mathbb{R}^{n \times n}$ con $\det A \neq 0$, es un abierto de $\mathbb{R}^{n \times n}$ ya que $\det : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ es un polinomio, y por tanto continuo. Luego por ser un abierto de $\mathbb{R}^{n \times n}$, resulta que $\text{Gl}(n, \mathbb{R})$ es una variedad diferenciable.

Por otro lado, si $A = (a_{ij})$, $B = (b_{ij})$ en $\text{Gl}(n, \mathbb{R})$, se tiene que $(AB)_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$, luego la función producto $\text{Gl}(n, \mathbb{R}) \times \text{Gl}(n, \mathbb{R}) \rightarrow \text{Gl}(n, \mathbb{R})$ es una función diferenciable. Además, $A^{-1} = \frac{1}{\det A} \text{adj}(A)$ y por lo tanto la función inversa $\text{Gl}(n, \mathbb{R}) \rightarrow \text{Gl}(n, \mathbb{R})$ es diferenciable.

En conclusión, $\text{Gl}(n, \mathbb{R})$ es una variedad diferenciable, un grupo, y además las operaciones del grupo son diferenciables. Esto es lo que llamamos grupo de Lie.

Definición C.1 (Grupo de Lie). Sea G una variedad diferenciable. Entonces decimos que G es un *grupo de Lie* si

1. G es un grupo
2. las operaciones $G \times G \rightarrow G$, $(x, y) \mapsto xy$ y $G \rightarrow G$, $x \mapsto x^{-1}$ son diferenciables.

Definición C.2 (Morfismos de grupos de Lie). Sean H y G grupos de Lie. Un mapa $\phi : H \rightarrow G$ es un *morfismo de grupos de Lie* o simplemente un *morfismo* si es un morfismo de grupos diferenciable.

- Ejemplos C.3.**
1. $(\mathbb{R}^n, +)$, $(\mathbb{C}^n, +)$ son grupos de Lie.
 2. \mathbb{R}^* , \mathbb{C}^* con la operación producto son grupos de Lie.
 3. S^1 es un grupo de Lie. En efecto, $S^1 = \mathbb{R}/\mathbb{Z}$, al ser \mathbb{Z} subgrupo normal de \mathbb{R} , se tiene que S^1 es un grupo. Además por ser \mathbb{Z} discreto, S^1 es una variedad diferenciable y además las operaciones pasan diferenciablemente al cociente. Otra manera es considerar $S^1 \subset \mathbb{C}^*$ como subgrupo que lo veremos en seguida.
 4. G, H grupos de Lie, entonces $G \times H$ con la multiplicación coordinada a coordinada es un grupo de Lie. En particular tenemos que el toro $\mathbb{T}^n = S^1 \times \dots \times S^1$ es un grupo de Lie.

Definición C.4 (Subgrupo de Lie). Un *subgrupo de Lie* H de un grupo de Lie G es un subgrupo de G que además es un grupo de Lie con su estructura diferenciable y que además es una variedad inmersa.

Ejemplo C.5. En $\mathbb{T}^2 = S^1 \times S^1$, donde S^1 lo pensamos como subconjunto de \mathbb{C}^* , consideremos $\gamma: \mathbb{R} \rightarrow \mathbb{T}^2$ dada por $\gamma(t) = (e^{2\pi it}, e^{2\pi itc})$ con $c \in \mathbb{R} \setminus \mathbb{Q}$, observar que γ es un morfismo de grupos. Luego $H = \text{Im } \gamma \subset \mathbb{T}^2$ es un subgrupo de $\mathbb{T}^2 \subset \mathbb{C}^* \times \mathbb{C}^*$. Sobre H inducimos la estructura diferenciable que hace $\gamma: \mathbb{R} \rightarrow H$ un difeo ($U \subset H$ abierto sii $\gamma^{-1}(U)$ abierto en \mathbb{R} y cartas locales $\varphi \circ \gamma^{-1}: U \rightarrow \mathbb{R}$ siendo φ carta local de \mathbb{R}). Luego, H es un grupo de Lie y por tanto un subgrupo de Lie de \mathbb{T}^2 . Observar que H es denso en \mathbb{T}^2 y por lo tanto no puede ser un encaje.

El siguiente resultado importante lo hacemos sin demostración.

Teorema C.6 (Cartan). *Sea $H < G$ un subgrupo de un grupo de Lie G que es cerrado como subconjunto. Entonces, H es una subvariedad de G y por lo tanto un subgrupo de Lie. En particular tiene la topología inducida.* \square

Con este resultado en cuestión podemos construir más ejemplos de grupos de Lie, considerando subgrupos cerrados de \mathbb{R}^n o $\text{Gl}(n, \mathbb{R})$.

- Ejemplos C.7.**
1. *Grupo lineal especial* $SL(n, \mathbb{R}) = \{A \in \text{Gl}(n, \mathbb{R}) : \det A = 1\}$
 2. *Grupo ortogonal* $O(n) = \{A \in \text{Gl}(n, \mathbb{R}) : A^T A = \text{Id}\}$

3. *Grupo unitario* $\mathcal{U}(n) = \{A \in \mathbb{G}l(n, \mathbb{C}) : A^*A = \text{Id}\}$

4. *Grupo ortogonal especial* $SO(n) = \{A \in \mathcal{O}(n) : \det A = 1\}$

5. *Grupo unitario especial* $SU(n) = \{A \in \mathcal{U}(n) : \det A = 1\}$

Observación C.8. Los grupos $\mathcal{O}(n), \mathcal{U}(n)$ pueden identificarse con los grupos de isometrías lineales de \mathbb{R}^n o \mathbb{C}^n respectivamente.

Dado $\langle \cdot, \cdot \rangle$ producto interno sobre \mathbb{R}^n , $\langle x, y \rangle = \sum x_i y_i$, tenemos que

$$\mathcal{O}(n) = \{A \in \mathbb{R}^{n \times n} : \langle Ax, Ay \rangle = \langle x, y \rangle, \forall x, y \in \mathbb{R}^n\}.$$

Análogo para $\mathcal{U}(n)$.

C.1. Espacio tangente a un grupo de Lie

A continuación veremos la importancia que tiene en un grupo de Lie el espacio tangente a la identidad. Para ver esto consideremos los siguientes mapas. Dado $a \in G$, se define la *traslación a izquierda* por a al mapa $L_a : G \rightarrow G$ dado por $L_a(g) = ag$. Análogamente sea $R_a : G \rightarrow G$ dado por $R_a(g) = ga$ la *traslación a derecha*. Observar que son difeomorfismos, ya que $(L_a)^{-1} = L_{a^{-1}}$. Estos mapas nos permiten movernos dentro de G . Por ejemplo, todo $a \in G$ puede ser trasladado a la identidad, y también sus vectores tangentes, pues $dL_{a^{-1}}(a) : T_a G \rightarrow T_e G$ es un isomorfismo.

Definición C.9 (Campos invariantes). Decimos que un campo vectorial $X \in \mathcal{X}(G)$ es *invariante a izquierda* si $X_{ag} = dL_a(g)X_g$ para todo $a, g \in G$.

Observación C.10. Un campo X invariante a izquierda queda determinado por su valor $X_e \in T_e G$. En efecto, $X_g = dL_g(e)X_e$. Además por ser la traslación diferenciable, todo campo invariante a izquierda es automáticamente diferenciable.

Definición C.11. \mathfrak{g} es el conjunto de *campos invariantes a izquierda*. Es un espacio vectorial con las operaciones habituales.

◇ **C.12.** Probar que si $X, Y \in \mathfrak{g}$, entonces $[X, Y] \in \mathfrak{g}$.

En este sentido el espacio vectorial \mathfrak{g} , dotado del corchete de Lie, es un *álgebra de Lie*.

Proposición C.13. *El mapa $\mathfrak{g} \rightarrow T_e G$, dado por $X \mapsto X_e$ es un isomorfismo.*

Demostración. Si $X_e = 0$, al ser invariante a izquierda se tiene, $X_g = dL_g(e)X_e = 0 \forall g \in G$, por lo tanto $X = 0$. Ahora, dado $v \in T_e G$, definimos el campo X^v dado por $X_g^v = dL_g(e)v$. Es claro que $X^v \in \mathfrak{g}$ y satisface $X_e^v = v$. \square

Este isomorfismo define un corchete de Lie sobre $T_e G$, a saber

$$[u, v] := [X^u, X^v](e).$$

Ejemplos C.14. 1. Es un ejercicio verificar que el álgebra de Lie asociada a $\mathbb{G}l(n, \mathbb{R})$, $\mathfrak{gl}(n, \mathbb{R})$ es $\mathbb{R}^{n \times n}$ con el corchete de Lie matricial $[X, Y] = XY - YX$.

2. Análogamente, $\mathfrak{sl}(n, \mathbb{R}) = \{X \in \mathbb{R}^{n \times n} : \text{tr}(X) = 0\}$ es el álgebra de Lie de $SL(n, \mathbb{R})$.

3. $\mathfrak{antisim}(n, \mathbb{R}) = \{X \in \mathbb{R}^{n \times n} : X^T = -X\}$ es el álgebra de Lie de $\mathcal{O}(n)$.

4. $\mathfrak{Skew}(n, \mathbb{C}) = \{X \in \mathbb{C}^{n \times n} : X^* = -X\}$ es el álgebra de Lie de $\mathcal{U}(n)$.

En todos los casos el corchete de Lie es corchete matricial.

C.2. Subgrupos a un parámetro

Definición C.15 (Subgrupo a un parámetro). Un *subgrupo a un parámetro* de un grupo de Lie G es un morfismo de $(\mathbb{R}, +)$ a G . Es decir, una curva diferenciable sobre G .

Ejemplos C.16. 1. $\phi(t) = e^t$ es un subgrupo a un parámetro de $(\mathbb{R}, +)$

2. $\phi(t) = e^{it}$ es un subgrupo a un parámetro de $S^1 = \mathcal{U}(1)$

3. $\phi(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$ es un subgrupo a un parámetro de $\mathcal{O}(2)$

4. $\phi(t) = e^{tA}$ es un subgrupo a un parámetro de $\mathbb{G}l(n, \mathbb{R})$

El resultado central de esta sección es el siguiente.

Teorema C.17. Sea G un grupo de Lie y \mathfrak{g} su álgebra de Lie. Entonces, $\forall X \in \mathfrak{g}$, existe un único subgrupo a un parámetro ϕ tal que $\phi(0) = e$ y $\dot{\phi}(0) = X_e$.

Demostración. Sea $X \in \mathfrak{g}$, luego, por existencia y unicidad de las soluciones de ecuaciones diferenciales ordinarias, existe $\phi : (-\varepsilon, \varepsilon) \rightarrow G$ tal que $\phi(0) = e$ y $\dot{\phi}(t) = X_{\phi(t)}$. Sean $s, t, s+t \in (a, b)$ siendo (a, b) un intervalo donde ϕ esté definida. Luego, si $\phi_1(t) = \phi(s+t)$ y $\phi_2(t) = \phi(s)\phi(t)$ resulta que $\phi_1(0) = \phi_2(0)$. Además, $\dot{\phi}_1(t) = \dot{\phi}(s+t) = X_{\phi(s+t)} = X_{\phi_1(t)}$, luego ϕ_1 es una curva integral de X . Por otro lado, $\dot{\phi}_2(t) = \frac{d}{dt}\phi(s)\phi(t) = \frac{d}{dt}L_{\phi(s)}(\phi(t)) = DL_{\phi(s)}(\phi(t))\dot{\phi}(t)$, entonces $\dot{\phi}_2(t) = DL_{\phi(s)}(\phi(t))X_{\phi(t)} = X_{\phi(s)\phi(t)} = X_{\phi_2(t)}$ y por lo tanto ϕ_2 es una curva integral de X . Dado que hay unicidad de curvas integrales, tenemos que $\phi_1(t) = \phi_2(t)$. Es decir, ϕ es un morfismo de grupos donde este definida. Veamos ahora que ϕ está definida para todo \mathbb{R} . Sea $\gamma(t) = \phi(\frac{\varepsilon}{2})\phi(\frac{t-\varepsilon}{2})$. Observar que γ está definida en $(-\frac{\varepsilon}{2}, \frac{3\varepsilon}{2})$ y que $\gamma(t) = \phi_2(t)$ si tomamos $s = \frac{\varepsilon}{2}$, por lo tanto $\gamma(0) = \phi(0)$ y $\dot{\gamma}(t) = X_{\gamma(t)}$. Luego γ extiende a ϕ en $(-\varepsilon, \frac{3\varepsilon}{2})$, trabajando por inducción concluimos que ϕ está definida para todo $t \in \mathbb{R}$. Falta probar que si $\phi : \mathbb{R} \rightarrow G$ es un subgrupo a un parámetro con $\dot{\phi}(t) = X_e$, entonces ϕ es la curva integral asociada. Sea $X \in \mathfrak{g}$ el campo invariante a izquierda asociado. Se tiene

$$\begin{aligned} \dot{\phi}(t) &= \frac{d}{dt}\Big|_{s=0} \phi(t+s) &= \frac{d}{dt}\Big|_{s=0} \phi(t)\phi(s) &= \frac{d}{dt}\Big|_{s=0} L_{\phi(t)}(\phi(s)) \\ &= DL_{\phi(t)}(\phi(0))\dot{\phi}(0) &= DL_{\phi(t)}(e)X_e = X(\phi(t)). \end{aligned}$$

Es decir, ϕ es una curva integral. \square

Definición C.18. Sean G un grupo de Lie, y \mathfrak{g} su álgebra de Lie. El *mapa exponencial* está definido por

$$\exp : \mathfrak{g} \rightarrow G, \exp(X) = \phi_X(1),$$

siendo ϕ_X el subgrupo a un parámetro definido por $X \in \mathfrak{g}$.

\diamond **C.19.** Probar que $\exp(tX) = \phi_X(t)$ para todo $t \in \mathbb{R}$.

Corolario C.20. Sea $X \in \mathfrak{g}$. La curva $\gamma(t) = \exp(tX)$ es el único morfismo en G con $\gamma'(0) = X$. Además, por ser ϕ_X un morfismo, resulta que

$$1. \exp((s+t)X) = \exp(sX)\exp(tX) \quad \forall s, t \in \mathbb{R}$$

$$2. \exp(tX)^{-1} = \exp(-tX) \quad \forall t \in \mathbb{R}$$

Observación C.21. El mapa exponencial $\exp : \mathfrak{g} \rightarrow G$ es un mapa diferenciable donde $d_0 \exp : T_0 \mathfrak{g} = \mathfrak{g} \rightarrow T_e G = \mathfrak{g}$. Utilizando que $\exp(tX) = \phi_X(t)$ resulta

$$d_0 \exp(X) = \left. \frac{d}{dt} \right|_{t=0} \exp(tX) = \left. \frac{d}{dt} \right|_{t=0} \phi_X(t) = \dot{\phi}_X(0) = X_e \equiv X.$$

Es decir, con las identificaciones anteriores, $d_0 \exp : \mathfrak{g} \rightarrow \mathfrak{g}$ es la identidad. Luego se concluye el siguiente.

Teorema C.22. *Existe un abierto U de $0 \in \mathfrak{g}$ tal que el mapa exponencial es un difeomorfismo de U con $\exp(U)$ abierto de e en G .*

Comentario C.23. En $\mathbb{G}l(n, \mathbb{R})$ tenemos $\mathfrak{g} = \mathbb{R}^{n \times n}$. Consideremos $Exp : \mathbb{R}^{n \times n} \rightarrow \mathbb{G}l(n, \mathbb{R})$ dado por $Exp(A) = \sum \frac{A^n}{n!}$, la exponencial de matrices. Observar que si definimos $\phi(t) = Exp(tA)$ tenemos un subgrupo a un parámetro con $\dot{\phi}(0) = A$. Por lo tanto por el corolario anterior tenemos que es el mapa exponencial.

C.3. Acciones de grupos

La definición axiomática de grupos es relativamente reciente. Históricamente los grupos surgieron como “grupos” de transformaciones de ciertos objetos. Por ejemplo este fue el punto de vista de Galois, quien considero el grupo de permutaciones de raíces de polinomios. En esta dirección es que se motiva la siguiente definición.

Definición C.24 (Acción de un grupo). Sea G un grupo y X un conjunto. Decimos que G actúa en X y escribimos $G \curvearrowright X$, si hay un mapa $\sigma : G \times X \rightarrow X$ (denotamos $g \cdot x := \sigma(g, x)$) tal que

1. $e \cdot x = x$
2. $g \cdot (h \cdot x) = (gh) \cdot x \quad \forall g, h \in G, x \in X.$

Si G es un grupo topológico y X un espacio topológico entonces se requiere que σ sea continua. Análogamente si G es un grupo de Lie y X una variedad diferenciable, se requiere que σ sea suave.

Observación C.25. Para cada $g \in G$, el mapa $\sigma_g : X \rightarrow X$ dado por $\sigma_g(x) = g \cdot x$, define un automorfismo de X . Esto resulta de que $\sigma_{g^{-1}} : X \rightarrow X$ satisface $\sigma_{g^{-1}} \circ \sigma_g = id_X$. En este sentido, tenemos un mapa $\rho : G \rightarrow \text{Aut}(X)$, $g \mapsto \sigma_g$. Cuando X es un espacio vectorial, ρ es una *representación* de G . Observar además que ρ es un morfismo de grupos.

Recíprocamente, un morfismo de grupos $\rho : G \rightarrow \text{Aut}(X)$ define una acción de G en X , a saber, $g \cdot x = \rho(g)(x)$.

Ejemplos C.26. 1. $\text{Gl}(n, \mathbb{R}) \curvearrowright \mathbb{R}^n$, $A \cdot x = Ax$

2. Si $H \subset \text{Gl}(n, \mathbb{R})$ es un subgrupo de Lie, entonces $\sigma_H : H \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ define una acción dado que $\sigma_H = \sigma \circ (\tau \times id_{\mathbb{R}^n})$, siendo τ la inclusión. En particular $\mathcal{O}(n) \curvearrowright \mathbb{R}^n$.

3. $\mathcal{O} \curvearrowright \text{Sym}(n)$, $U \cdot A = UAU^{-1}$

Definición C.27 (Órbita de un grupo). Dada $G \curvearrowright X$, se define la *órbita* de $x \in X$ como el subconjunto $G \cdot x = \{g \cdot x : g \in G\}$. Decimos que $x \in X$ es un *punto fijo* si $G \cdot x = x$.

La acción se dice *transitiva* si existe un $x \in X$ tal que $G \cdot x = X$.

Ejemplos C.28. 1. Con $\text{Gl}(n, \mathbb{R}) \curvearrowright \mathbb{R}^n$, 0 es punto fijo, la acción es transitiva en $\mathbb{R}^n \setminus \{0\}$.

2. Con $\mathcal{O}(n) \curvearrowright \mathbb{R}^n$, tiene como órbitas esferas centradas en 0 .

Cociente por la acción

La gran mayoría de las variedades tratadas en este curso resultan de cocientar un espacio por la acción de un grupo de Lie.

Una acción de un grupo de Lie G en una variedad M induce una relación de equivalencia en M , a saber,

$$x \sim y \Leftrightarrow \exists g \in G : y = g \cdot x.$$

En particular, la clase $[x]$ de $x \in M$ coincide con la órbita de x . Denotamos por M/G al espacio cociente, es decir al espacio de órbitas

$$M/G = \{G \cdot x : x \in M\}.$$

En M/G inducimos la topología cociente (la más fina que hace $\pi : M \rightarrow M/G$ continua).

Lema C.29. *La proyección canónica $\pi : M \rightarrow M/G$ es abierta.*

Demostración. Sea $U \subset M$ abierto, entonces $\sigma_g(U)$ es un abierto en M para todo $g \in G$. Luego

$$\pi^{-1}(\pi(U)) = G \cdot U = \{g \cdot u : g \in G, u \in U\} = \bigcup_{g \in G} \sigma_g(U)$$

que es una unión de abiertos. □

Consideremos ahora $H < G$ subgrupo de Lie. Observar que $H \curvearrowright G$ a izquierda por $h \cdot g = gh^{-1}$. Por lo tanto el espacio cociente $G/H = \{gH : g \in G\}$.

Teorema C.30. *$\pi : G \rightarrow G/H$ es continua y abierta. Además, G/H es Hausdorff si y sólo si H es cerrado.*

Demostración. La continuidad y que es abierta ya lo probamos. Resta probar la segunda afirmación, y utilizando el ejercicio A.36 basta probar que H es cerrado si y sólo si $R = \{(x, y) \in G \times G : x \sim y\}$ es cerrado en $G \times G$. Sea $F : G \times G \rightarrow G$ dado por $F(x, y) = y^{-1}x$. Observar que F es suave por ser G un grupo de Lie. Luego

$$F^{-1}(H) = \{(x, y) \in G \times G : y^{-1}x \in H\} = \{(x, y) \in G \times G : x \in yH\} = R.$$

Como F es continua, resulta que si H es cerrado, R también. Recíprocamente, si G/H es Hausdorff, R es cerrado. En particular los puntos son cerrados. Por lo tanto $H = \pi^{-1}(eH)$ es cerrado. □

C.4. Espacios homogéneos

En lo que sigue G es un grupo de Lie, M es una variedad diferenciable.

Definición C.31 (Espacio homogéneo). M se dice *espacio homogéneo* de G si G actúa en M transitivamente.

Ejemplos C.32. 1. S^{n-1} es un espacio homogéneo de $O(n)$

2. $\mathbb{R}^n \setminus \{0\}$ es un espacio homogéneo de $\text{Gl}(n, \mathbb{R})$
3. $S^{2n-1} \subset \mathbb{C}^n$ es un espacio homogéneo de $\mathcal{U}(n)$
4. $\mathbb{R}P^n$ es un espacio homogéneo de $\mathcal{O}(n+1)$
5. $\text{Grass}_{\mathbb{R}}(k, n)$ es un espacio homogéneo de $\mathcal{O}(n)$
6. $\text{Flag}(n, \mathbb{C})$ es un espacio homogéneo de $\mathcal{U}(n)$

Definición C.33 (Grupo de isotropía). Sean $G \curvearrowright M$ y $x \in M$. Se define el *grupo de isotropía* de x (o el estabilizador de x) como $G_x := \{g \in G : g \cdot x = x\}$.

Observación C.34. Consideremos la acción de G en M desde el punto de vista conjuntista. Dado $x_0 \in M$ sea $F : G \rightarrow M$ dada por $F(g) = g \cdot x_0$. Si $G \curvearrowright M$ es transitiva, resulta F sobreyectiva. Además, $F^{-1}(x_0) = H$ siendo H el grupo de isotropía de x_0 . Observar que F es equivariante por la acción de G , es decir $g \cdot F(g') = F(gg')$, por lo tanto $F^{-1}(x) = g_x H$ siendo $g_x \in G$ tal que $g_x \cdot x_0 = x$.

Luego el espacio cociente inducido por F coincide con el cociente inducido por la acción $H \curvearrowright G$, que denotamos G/H . Entonces $\tilde{F} : G/H \rightarrow M$ es un mapa biyectivo y equivariante por la acción de G . Por lo tanto concluimos que desde el punto de vista conjuntista G/H y M son lo mismo y las acciones $\sigma : G \times M \rightarrow M$ y $\lambda : G \times G/H \rightarrow G/H$ son equivalentes.

Teorema C.35. Si G es un grupo de Lie y H es un subgrupo cerrado de G , entonces existe una única estructura diferenciable en G/H tal que la proyección $\pi : G \rightarrow G/H$ es una submersión. Además, $\dim G/H = \dim G - \dim H$, y $\lambda : G \times G/H \rightarrow G/H$ es suave.

Observación C.36. Este resultado tiene una aplicación muy importante, a saber, dota de estructura diferenciable a conjuntos no tan triviales (como veremos en los ejemplos más adelante). El método es relativamente sencillo:

Sea G un grupo de Lie que actúa de manera transitiva sobre un conjunto X , de tal manera que existe algún $x \in X$ con subgrupo de isotropía G_x cerrado. Luego la biyección entre G/G_x y X permite dotar a X con una única estructura diferenciable que hace la acción suave.

Cuando el conjunto ya tiene una estructura diferenciable, resulta el siguiente.

Teorema C.37. Sean $G \curvearrowright M$ transitiva y H el grupo de isotropía de un $x \in M$. Entonces,

1. H es un subgrupo cerrado de G .
2. El mapa $\tilde{F} : G/H \rightarrow M$ dado por $\tilde{F}(gH) = g \cdot x$ es un difeomorfismo.
3. $\dim G/H = \dim G - \dim H$.

Con este resultado en mente podemos redefinir un espacio homogéneo.

Definición C.38 (Espacio homogéneo). Un *espacio homogéneo* es una variedad M con una acción transitiva de un grupo de Lie. Equivalentemente es una variedad de la forma G/H donde G es un grupo de Lie y H un subgrupo cerrado.

Ejemplos C.39. 1. Un grupo de Lie G es trivialmente un espacio homogéneo tomando la acción trivial y $\{e\}$ como subgrupo.

2. (La esfera S^n) El grupo ortogonal $\mathcal{O}(n+1)$ actúa en $S^n \subset \mathbb{R}^{n+1}$ con la restricción de la acción de $\mathbb{G}l(n+1, \mathbb{R})$ en \mathbb{R}^{n+1} (ver ejemplo). Además la acción es transitiva. El grupo de isotropía de $(1, 0, \dots, 0) = e_1 \in \mathbb{R}^{n+1}$ son las matrices de la forma $\begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix} \in \mathcal{O}(n+1)$, donde $Q \in \mathcal{O}(n)$. Identificando este subgrupo con $\mathcal{O}(n)$, obtenemos que $\mathcal{O}(n+1)/\mathcal{O}(n)$ es difeomorfo a S^n . Análogamente para $SO(n+1)$, tenemos que $SO(n+1)/SO(n)$ es difeomorfo a S^n . En particular S^n es conexo.
3. Análogamente al caso anterior, pero para el caso complejo, resulta que la esfera real $S^{2n+1} \subset \mathbb{C}^{n+1}$, se puede describir como espacio homogéneo $S^{2n+1} = \mathcal{U}(n+1)/\mathcal{U}(n)$.
4. (Espacio proyectivo $\mathbb{R}P^n$ y $\mathbb{C}P^n$) El grupo $\mathcal{O}(n+1)$ actúa de manera natural sobre $\mathbb{R}P^n$ ($l \in \mathbb{R}P^n \mapsto Ql \in \mathbb{R}P^n$). El subgrupo de isotropía de $(1 : 0 : \dots : 0)$ son las matrices de la forma $\begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}$ ó $\begin{pmatrix} -1 & 0 \\ 0 & Q' \end{pmatrix}$ con $Q, Q' \in \mathcal{O}(n)$. Observando que $\mathcal{O}(1) = \{1, -1\}$, obtenemos $\mathbb{R}P^n = \mathcal{O}(n+1)/\mathcal{O}(n) \times \mathcal{O}(1)$. Análogamente, $\mathbb{C}P^n = \mathcal{U}(n+1)/\mathcal{U}(n) \times \mathcal{U}(1)$.

5. (Grassmaniana) El grupo $\mathcal{O}(n)$ actúa de manera natural sobre subespacios de \mathbb{R}^n . Luego $\mathcal{O}(n) \curvearrowright \text{Grass}_{\mathbb{R}}(k, n)$. Además, esta acción es transitiva. Sea V el subespacio generado por $\{e_1, \dots, e_k\} \subset \mathbb{R}^n$. El subgrupo de isotropía asociado a V está dado por matrices de la forma $\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$ con $A \in \mathcal{O}(k)$ y $B \in \mathcal{O}(n-k)$. Luego $\text{Grass}_{\mathbb{R}}(k, n) = \mathcal{O}(n)/\mathcal{O}(k) \times \mathcal{O}(n-k)$. Observar que de esta manera es clara la identificación entre $\text{Grass}_{\mathbb{R}}(k, n)$ y $\text{Grass}_{\mathbb{R}}(n-k, n)$. Análogamente se prueba $\text{Grass}_{\mathbb{C}}(k, n) = \mathcal{U}(n)/\mathcal{U}(k) \times \mathcal{U}(n-k)$.
6. (Variedad de Stiefel) El grupo $\mathcal{O}(n)$ actúa sobre $St(k, n)$ de manera natural y es transitiva. El subconjunto de isotropía de $X = (e_1, \dots, e_k) \in St(k, n)$ es el subconjunto de las matrices de la forma $\begin{pmatrix} \text{Id} & 0 \\ 0 & Q \end{pmatrix}$ con $Q \in \mathcal{O}(n-k)$. Luego $St(k, n) = \mathcal{O}(n)/\mathcal{O}(n-k)$. Análogamente si consideramos la versión compleja $St_{\mathbb{C}}(k, n) = \{X \in \mathbb{C}^{n \times k} : X^*X = \text{Id}\}$, resulta $St_{\mathbb{C}}(k, n) = \mathcal{U}(n)/\mathcal{U}(n-k)$.
7. (Espacio bandera) Como ya vimos en la Sección 2.22, el espacio bandera $\text{Flag}(\mathbb{C}^n)$ formado por $F = (F_0, \dots, F_n)$ siendo F_i subespacios de dimensión i tales que $F_i \subset F_{i+1}$, puede identificarse por el cociente $\mathbb{G}\text{L}(n, \mathbb{C})/\mathcal{R}_n(\mathbb{C})$ ó $\mathcal{U}(n)/\mathcal{U}^1(n)$ de las acciones de $\mathbb{G}\text{L}(n, \mathbb{C})$ y $\mathcal{U}(n)$ en $\text{Flag}(\mathbb{C}^n)$ respectivamente.
8. (Banderas generalizadas) Sean n_1, \dots, n_r enteros positivos tales que $n_1 + \dots + n_r = n$ y $\text{Flag}_{\mathbb{K}}(n_1, \dots, n_r)$ el conjunto de banderas parciales $F = (F_1, \dots, F_r)$ tales que $\dim F_i = n_1 + \dots + n_i$, y $F_i \subset F_{i+1}$ para $i = 1, \dots, r-1$. Motivado por el ejemplo anterior, $\text{Flag}_{\mathbb{K}}(n_1, \dots, n_r)$ es el espacio homogéneo $\mathcal{O}(n)/\mathcal{O}(n_1) \times \dots \times \mathcal{O}(n_r)$ y $\mathcal{U}(n)/\mathcal{U}(n_1) \times \dots \times \mathcal{U}(n_r)$ para $\mathbb{K} = \mathbb{R}$ ó \mathbb{C} respectivamente.

Observación C.40. Una forma de verificar que las estructuras diferenciables de los ejemplos anteriores coinciden, es probar que la acción del grupo en la variedad es suave.

C.5. Métricas Riemannianas en grupos de Lie

Todo grupo de Lie posee métricas naturales que preservan cierta invarianza. Estas son las métricas invariantes a izquierda o derecha, o bi-invariantes.

Definición C.41 (Métricas bi-invariantes). Sea G un grupo de Lie. Una *métrica Riemanniana invariante a izquierda* en G es una métrica Riemanniana tal que las traslaciones a izquierda son isometrías. Análogo para *métricas Riemannianas invariantes a derecha* ó *bi-invariantes*.

Construir una métrica invariante a izquierda es sencillo. Sea $\langle \cdot, \cdot \rangle_e$ un producto interno en \mathfrak{g} (o $T_e G$) álgebra de Lie de G . Luego mediante la traslación a izquierda $L_g : G \rightarrow G$ podemos "pull-backear" la métrica en cualquier tangente $T_g G$:

$$\langle u, v \rangle_g := \langle DL_g(e)^{-1}u, DL_g(e)^{-1}v \rangle_e, \quad u, v \in T_g G, \quad g \in G \quad (\text{C.1})$$

Dado que esta métrica es constante sobre campos invariantes a izquierda, es decir $\langle X_g, Y_g \rangle_g = \langle X_e, Y_e \rangle_e$, y de que todo campo diferenciable en G es combinación lineal de campos invariantes a izquierda con coeficientes en $\mathcal{C}^\infty(G)$, resulta que los productos internos definidos en C.1 son una métrica Riemanniana. Equivalentemente se puede definir una métrica invariante imponiendo que ciertos campos invariantes a izquierda sean base ortonormal.

Proposición C.42. *Las métricas invariantes a izquierda están en correspondencia con productos internos en \mathfrak{g} .*

Observación C.43. Como composición de isometrías es una isometría, resulta que $DL_g(h) : T_h G \rightarrow T_{gh} G$ son isometrías para cualesquiera $g, h \in G$.

C.5.1. Métricas bi-invariantes

La existencia de métricas bi-invariantes en un grupo de Lie es un poco más complicada. Para estudiar su existencia es importante considerar el automorfismo $I_x : G \rightarrow G$ dada por $I_x(g) = xgx^{-1}$. El mapa I_x se llama *automorfismo interno* de G , y coincide con $I_x = L_x \circ R_{x^{-1}} = R_{x^{-1}} \circ L_x$.

Definición C.44 (Representación adjunta). La *representación adjunta* de G es el morfismo

$$Ad : G \rightarrow \text{Aut}(\mathfrak{g}), Ad(g) = DI_g(e).$$

Observación C.45. Dado que $I_x \circ I_y = I_{xy}$, resulta que $Ad(x)Ad(y) = Ad(xy)$. De la igualdad $I_x = L_x \circ R_{x^{-1}}$ resulta

$$Ad(x) : \mathfrak{g} \rightarrow \mathfrak{g}, Ad(x) = DL_x(x^{-1})DR_{x^{-1}}(e)$$

y por lo tanto $\langle \cdot, \cdot \rangle_e$ es una métrica bi-invariante si y sólo si $Ad(x)$ es una isometría para todo $x \in G$. En este caso la métrica se dice *Ad-invariante*.

En el curso asumiremos el siguiente resultado sin demostración.

Teorema C.46. *Todo grupo de Lie compacto está dotado de una métrica Riemanniana bi-invariante.* □

Ejemplo C.47. (Métrica bi-invariante en $\mathcal{O}(n), \mathcal{U}(n)$) Sea $\langle \cdot, \cdot \rangle_F$ el producto interno en el espacio de matrices $n \times n$, es decir $\langle A, B \rangle_F = \text{tr}(A^T B)$. En $\mathcal{O}(n)$ podemos definir la métrica invariante a izquierda que resulta de inducir $\langle \cdot, \cdot \rangle_{\text{Id}}$ en $T_{\text{Id}}\mathcal{O}(n)$. Si $U \in \mathcal{O}(n)$, se tiene $DL_U(X)\dot{X} = U\dot{X}$ para todo $\dot{X} \in T_X\mathcal{O}(n)$.

Definimos

$$\begin{aligned} \langle \dot{X}, \dot{Y} \rangle_X &:= \text{tr}((DL_X(\text{Id})^{-1}\dot{X})^T(DL_X(\text{Id})^{-1}\dot{Y})) \\ &= \text{tr}((X^{-1}\dot{X})^T(X^{-1}\dot{Y})) \\ &= \text{tr}(\dot{X}^T(X^{-1})TX^{-1}\dot{Y}) \\ &= \text{tr}(\dot{X}^T\dot{Y}). \end{aligned}$$

En particular, la métrica invariante a izquierda inducida coincide con la métrica invariante inducida por el ambiente.

Además es fácil ver que la métrica es bi-invariante.

Análogo para $\mathcal{U}(n)$, con $\langle A, B \rangle_{\text{Id}} = \text{tr}(B^*A)$.

C.6. Métricas Riemannianas en espacios homogéneos

Sea G un grupo de Lie, H un subgrupo cerrado de G . El espacio homogéneo G/H puede dotarse de una estructura diferenciable de manera tal que la proyección

$\pi : G \rightarrow G/H$ sea una submersión y la acción $\lambda : G \times G/H \rightarrow G/H$ suave.

Tomando diferenciales se tiene que $D\pi(e) : \mathfrak{g} = T_e G \rightarrow T_e G/H$ es un mapa lineal sobreyectivo.

Afirmación: $\text{Ker } D\pi(e) = \mathfrak{h}$, siendo \mathfrak{h} el álgebra de H . Esto resulta de que $\pi|_H$ es constante igual a eH , siendo H subvariedad de G , y que $\pi(g) = \pi(e)$ sii $g \in H$.

Luego el tangente $T_e G/H$ puede identificarse con $\mathfrak{g}/\mathfrak{h}$. Si G posee una métrica podemos identificar $T_e G/H$ con \mathfrak{m} subespacio de \mathfrak{g} siendo $\mathfrak{m} = \mathfrak{h}^\perp$.

¿ Cómo podemos definir una estructura Riemanniana sobre G/H ?

Una forma posible es definir el producto interno en $T_g G/H$ inducido por π , i.e. $\langle \xi, \zeta \rangle_{gH} := \langle D\pi(g)^{-1}|_{H_{\pi,g}} \xi, D\pi(g)^{-1}|_{H_{\pi,g}} \zeta \rangle_g$ siendo $H_{\pi,g} = \text{Ker}(D\pi(g))^\perp$ es espacio horizontal. En particular, $H_{\pi,e} = \mathfrak{m}$, $H_{\pi,g} = DL_g(e)\mathfrak{m}$.

¿ La definición anterior es independiente del representante?

Sea $g' = gh$, con $h \in H$. Luego para que la definición anterior tenga sentido es necesario que

$$\langle D\pi(g')^{-1}|_{H_{\pi,g'}} \xi, D\pi(g')^{-1}|_{H_{\pi,g'}} \zeta \rangle = \langle D\pi(g)^{-1}|_{H_{\pi,g}} \xi, D\pi(g)^{-1}|_{H_{\pi,g}} \zeta \rangle.$$

Dado que $\pi \circ R_h = \pi$, se tiene

$$D\pi(g')^{-1}|_{H_{\pi,g'}} \xi = DR_h(g) \circ D\pi(g)^{-1}|_{H_{\pi,g}} \xi.$$

Por lo tanto se requiere que

$$\langle DR_h(g) D\pi(g)^{-1}|_{H_{\pi,g}} \xi, DR_h(g) D\pi(g)^{-1}|_{H_{\pi,g}} \zeta \rangle = \langle D\pi(g)^{-1}|_{H_{\pi,g}} \xi, D\pi(g)^{-1}|_{H_{\pi,g}} \zeta \rangle.$$

Estas discusiones motivan y prueban el siguiente resultado.

Teorema C.48. *Sea G un grupo de Lie y H un subgrupo cerrado. Si $\langle \cdot, \cdot \rangle$ es una métrica Riemanniana invariante a izquierda en G y además es invariante a derecha por H , entonces en G/H puede inducirse una métrica Riemanniana tal que $\pi : G \rightarrow G/H$ es una submersión Riemanniana. Además la acción a izquierda de G en G/H deja la métrica invariante. \square*

Ejemplo C.49. Dado que $\mathcal{O}(n)$ y $\mathcal{U}(n)$ poseen métricas Riemannianas bi-invariantes resulta que esferas, proyectivos, grassmanianas, variedad de stiefel, espacios bandera pueden dotarse de manera natural con estructura Riemanniana que son invariantes por la acción del grupo base.

Conexiones

D.1. Introducción

El método de Newton es un método que sirve para hallar los ceros de una función $f : \mathbb{R}^n \rightarrow \mathbb{R}$, por lo que también puede ser utilizado para hallar los puntos críticos de una función haciendo uso de la información de segundo orden de la misma, concretamente:

$$\frac{\partial(\text{grad}(f))}{\partial v},$$

siendo $\frac{\partial}{\partial v}$ la derivada direccional en la dirección del vector $v \in \mathbb{R}^n$.

Cuando el método resulta convergente, la convergencia es superlineal. Para poder generalizar el método de Newton a una función $f : M \rightarrow \mathbb{R}$ donde M es una variedad, debemos introducir el concepto de *conexión afín* que generaliza el de derivada direccional de un campo respecto de un vector.

D.2. Conexiones afines

Sea M una variedad, $C^\infty(M)$ el conjunto de funciones a \mathbb{R} diferenciables y $\mathfrak{X}(M)$ el conjunto de campos tangentes en M .

Definición D.1 (Conexión). La función $\nabla : \mathfrak{X}(M) \times \mathfrak{X}(M) \rightarrow \mathfrak{X}(M)$ es una *conexión* si verifica las siguientes propiedades. Utilizaremos la notación siguiente $\nabla(X, Y) = \nabla_X Y$.

1. $C^\infty(M)$ linealidad en X

$$\nabla_{fX+gY}Z = f\nabla_X Z + g\nabla_Y Z.$$

2. \mathbb{R} linealidad en Y

$$\nabla_X aY + bZ = a\nabla_X Y + b\nabla_X Z.$$

3. Regla de Leibniz

$$\nabla_X fY = X(f)Y + f\nabla_X Y.$$

Usaremos indistintamente $X(p)$ ó X_p .

Observación D.2. En \mathbb{R}^n la derivada direccional define una conexión afín.

Sean X, Y campos en \mathbb{R}^n , es decir, $X, Y : \mathbb{R}^n \rightarrow \mathbb{R}^n$ diferenciables, la función definida como $(\nabla_X Y)_p = \frac{\partial Y}{\partial X_p}(p)$ verifica las tres propiedades de la definición. Es fácil ver que las propiedades de linealidad se verifican. Probemos que se verifica la regla de Leibniz:

$$\begin{aligned} (\nabla_X fY)_p &= \lim_{t \rightarrow 0} \frac{1}{t} [(fY)(p+tX_p) - (fY)(p)] \\ &= \lim_{t \rightarrow 0} \frac{1}{t} [(fY)(p+tX_p) - f(p+tX_p)Y(p) + f(p+tX_p)Y(p) - (fY)(p)] \\ &= \lim_{t \rightarrow 0} \frac{1}{t} [f(p+tX_p)(Y(p+tX_p) - Y(p)) + Y(p)(f(p+tX_p) - f(p))] \\ &= f(p) \frac{\partial Y}{\partial X_p}(p) + Y(p) \frac{\partial f}{\partial X_p}(p) = (f\nabla_X Y + YX(f))_p \end{aligned}$$

dado que $\frac{\partial f}{\partial X_p}(p) = (X(f))(p)$.

La anterior se denomina *conexión euclídeana canónica*, o simplemente *conexión canónica*.

Proposición D.3. *Toda variedad admite infinitas conexiones afines.* □

A continuación daremos una idea de la prueba. Para tal fin, utilizaremos las propiedades de la conexión para poder dar una expresión que nos permita utilizar como definición.

Consideremos en primer lugar el caso $M = \mathbb{R}^n$. Sea $\{e_1, \dots, e_n\}$ la base canónica, y consideremos los campos vectoriales $X = \sum_{i=1}^n X^i e_i$ y $Y = \sum_{j=1}^n Y^j e_j$. Por propiedades de conexión tenemos

$$\begin{aligned} \nabla_X Y &= \nabla_{\sum_{i=1}^n X^i e_i} \sum_{j=1}^n Y^j e_j = \sum_{i=1}^n X^i \sum_{j=1}^n \nabla_{e_i} Y^j e_j \\ &= \sum_{i=1}^n X^i \sum_{j=1}^n e_i(Y^j) e_j + Y^j \nabla_{e_i} e_j = \sum_{i,j=1}^n X^i (e_i(Y^j) e_j + Y^j \nabla_{e_i} e_j) \end{aligned}$$

De donde se concluye

$$\nabla_X Y = \sum_{i,j=1}^n X^i \left(\frac{\partial(Y^j \circ \phi)}{\partial x_i} e_j + Y^j \nabla_{e_i} e_j \right). \quad (\text{D.1})$$

Luego para que quede bien definida la conexión es necesario explicitar el campo vectorial $\nabla_{e_i} e_j$. Luego

$$\nabla_{e_i} e_j = \sum_{k=1}^n \Gamma_{ij}^{(k)} e_k, \quad (i, j = 1, \dots, n). \quad (\text{D.2})$$

Los símbolos $\Gamma_{ij}^{(k)}$ se denominan *símbolos de Christoffel* y definen la conexión.

De esta manera podemos concluir que \mathbb{R}^n puede ser dotado de infinitas conexiones: alcanza con especificar las n^3 funciones C^∞ , dadas por los símbolos de Christoffel.

Para probar el mismo resultado sobre una variedad se procede de igual manera tomando una carta local $\phi = (x_1, \dots, x_n) : V_p \subset M \rightarrow U \subset \mathbb{R}^n$, donde $\{D_1, \dots, D_n\}$ es la base canónica de $TV_p \subset TM$. Análogamente a lo anterior, si $X = \sum_{i=1}^n X^i D_i$ e $Y = \sum_{j=1}^n Y^j D_j$, utilizando las propiedades de conexión podemos concluir

$$\nabla_X Y = \sum_{i,j=1}^n X^i \left(\frac{\partial(Y^j \circ \phi)}{\partial x_i} D_j + Y^j \nabla_{D_i} D_j \right).$$

Luego eligiendo los símbolos de Christoffel

$$\nabla_{D_i} D_j = \sum_{k=1}^n \Gamma_{ij}^{(k)} D_k, \quad (i, j = 1, \dots, n),$$

obtenemos que localmente existen infinitas conexiones posibles. Para terminar la prueba es necesario probar que esta construcción es compatible con otra elección de cartas locales.

Observación D.4. También es posible dar una forma matricial a la expresión de la conexión.

$$\begin{aligned}
 \nabla_X Y &= \sum_{i,j=1}^n X^i \left(\frac{\partial(Y^j \circ \phi)}{\partial x_i} D_j + Y^j \nabla_{D_i} D_j \right) \\
 &= \sum_{j=1}^n \left(\sum_{i=1}^n X^i \frac{\partial(Y^j \circ \phi)}{\partial x_i} \right) D_j + \sum_{i,j,k=1}^n X^i Y^j \Gamma_{ij}^{(k)} D_k \\
 &= \sum_{j=1}^n \left(\frac{\partial(Y^j \circ \phi)}{\partial X} \right) D_j + \sum_{i,j,k=1}^n X^i Y^j \Gamma_{ij}^{(k)} D_k \\
 &= \frac{\partial(Y \circ \phi)}{\partial X} + \sum_{i,j,k=1}^n X^i Y^j \Gamma_{ij}^{(k)} D_k
 \end{aligned}$$

Tenemos que el campo $\nabla_X Y$ se compone por

- la derivada direccional (en coordenadas locales) de Y respecto del vector (X_1, \dots, X_n) ,
- un sumando que depende tanto de la conexión particular que estemos considerando como de los valores puntuales de los campos X e Y .
- Los símbolos de Christoffel dependen de la carta utilizada.
- En \mathbb{R}^n , la conexión canónica corresponde a considerar $\Gamma_{ij}^{(k)} = 0$ para todo $i, j, k = 1, \dots, n$.

Observación D.5. 1. $(\nabla_X Y)_p$ depende puntualmente de X , por lo cual podemos pensar en una función $(\nabla)_p : T_p M \times \mathfrak{X}(M) \rightarrow \mathfrak{X}(M)$

2. $(\nabla_X Y)_p$ depende localmente de Y . En particular, si $Y = Z$ en un entorno de p entonces se cumple que $(\nabla_X Y)_p = (\nabla_X Z)_p$.

3. Si ∇ y $\tilde{\nabla}$ son dos conexiones se cumple que $(\nabla_X Y - \tilde{\nabla}_X Y)_p$ depende puntualmente de Y , en efecto,

$$\nabla_X Y - \tilde{\nabla}_X Y = \sum_{i,j=1}^n \left[X^i \left(\frac{\partial(Y^j \circ \phi)}{\partial x_i} D_j + Y^j \nabla_{D_i} D_j \right) - X^i \left(\frac{\partial(Y^j \circ \phi)}{\partial x_i} D_j + Y^j \tilde{\nabla}_{D_i} D_j \right) \right] =$$

$$= \sum_{i,j=1}^n X^i Y^j (\nabla_{D_i} D_j - \tilde{\nabla}_{D_i} D_j).$$

En particular, si $Y_p = 0$ obtenemos que

$$(\nabla_X Y)_p = (\tilde{\nabla}_X Y)_p.$$

En particular, si $f \in C^\infty(M)$, siendo M una variedad diferenciable, entonces $\nabla_X \text{grad } f$ no depende de la elección de la conexión elegida. (Cf. Sección B.6 y Sección D.7.)

D.3. Conexión Riemanniana

Si bien existen infinitas conexiones en una variedad riemanniana tenemos una distinguida que satisface ciertas condiciones adicionales, y cuando $M = \mathbb{R}^n$ coincide con la conexión euclideana canónica.

Definición D.6 (Conexión simétrica). Decimos que una conexión ∇ es *simétrica* si $\Gamma_{ij}^{(k)} = \Gamma_{ji}^{(k)}$ para todo i, j, k .

Observación D.7. Si bien en principio esta definición dependería de la carta escogida, sucede que es equivalente a pedir que $\nabla_X Y - \nabla_Y X = [X, Y]$ lo que muestra la independencia de la carta escogida. En efecto, si la conexión es simétrica se cumple que

$$\begin{aligned} \nabla_X Y - \nabla_Y X &= \sum_{i,j=1}^n X^i (D_i(Y^j) D_j + Y^j \nabla_{D_i} D_j) - \left(\sum_{i,j=1}^n Y^i (D_i(X^j) D_j + X^j \nabla_{D_i} D_j) \right) = \\ &= \sum_{i,j=1}^n X^i D_i(Y^j) D_j - \left(\sum_{i,j=1}^n Y^i D_i(X^j) D_j \right) = [X, Y]. \end{aligned}$$

Recíprocamente, si $\nabla_{D_i} D_j - \nabla_{D_j} D_i = [D_i, D_j]$, entonces por ser $[D_i, D_j] = 0$ (ver en coordenadas locales), resulta que $\nabla_{D_i} D_j = \nabla_{D_j} D_i$ y por lo tanto ∇ es simétrica.

Definición D.8 (Conexión Riemanniana). Decimos que una conexión ∇ es *compatible con la métrica* si

$$Z(\langle X, Y \rangle) = \langle \nabla_Z X, Y \rangle + \langle X, \nabla_Z Y \rangle.$$

Asumamos el siguiente resultado.

Teorema D.9 (Levi-Civita). *Sea M variedad Riemanniana, entonces existe una única conexión Riemanniana simétrica y compatible con la métrica. Dicha conexión se denomina conexión de Levi-Civita o Riemanniana.* \square

Observación D.10. La conexión de Levi-Civita queda caracterizada por la fórmula de Koszul:

$$2\langle \nabla_X Y, Z \rangle = X\langle Y, Z \rangle + Y\langle Z, X \rangle - Z\langle X, Y \rangle - \langle X, [Y, Z] \rangle + \langle Y, [Z, X] \rangle + \langle Z, [X, Y] \rangle$$

Demostración. Veamos que la conexión de Levi-Civita verifica la fórmula de Koszul

$$\begin{aligned} X(\langle Y, Z \rangle) + Y(\langle Z, X \rangle) - Z(\langle X, Y \rangle) - \langle X, [Y, Z] \rangle + \langle Y, [Z, X] \rangle + \langle Z, [X, Y] \rangle &= \\ \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle + \langle \nabla_Y Z, X \rangle + \langle Z, \nabla_Y X \rangle - \langle \nabla_Z X, Y \rangle - \langle X, \nabla_Z Y \rangle &= \\ -\langle X, \nabla_Y Z - \nabla_Z Y \rangle + \langle Y, \nabla_Z X - \nabla_X Z \rangle + \langle Z, \nabla_X Y - \nabla_Y X \rangle &= \\ 2\langle \nabla_X Y, Z \rangle. & \end{aligned}$$

Recíprocamente, veamos que si una conexión verifica la fórmula de Koszul, entonces se cumple tanto la propiedad de simetría como de compatibilidad respecto de la métrica. Veamos en primer lugar que verifica la propiedad de compatibilidad respecto de la métrica.

$$\begin{aligned} 2[\langle \nabla_Z X, Y \rangle + \langle X, \nabla_Z Y \rangle] &= \\ Z\langle X, Y \rangle + X\langle Y, Z \rangle - Y\langle Z, X \rangle - \langle Z, [X, Y] \rangle + \langle X, [Y, Z] \rangle + \langle Y, [Z, X] \rangle + Z\langle Y, X \rangle &= \\ + Y\langle X, Z \rangle - X\langle Z, Y \rangle - \langle Z, [Y, X] \rangle + \langle Y, [X, Z] \rangle + \langle X, [Z, Y] \rangle &= \\ = 2\langle Z, [X, Y] \rangle. & \end{aligned}$$

Veamos en ahora que verifica la propiedad de simetría, para todo campo Z se tiene que

$$\begin{aligned} 2(\langle \nabla_X Y, Z \rangle - \langle \nabla_Y X, Z \rangle) &= \\ X\langle Y, Z \rangle + Y\langle Z, X \rangle - Z\langle X, Y \rangle - \langle X, [Y, Z] \rangle + \langle Y, [Z, X] \rangle + \langle Z, [X, Y] \rangle &= \\ -Y\langle X, Z \rangle - X\langle Z, Y \rangle + Z\langle Y, X \rangle + \langle Y, [X, Z] \rangle - \langle X, [Z, Y] \rangle - \langle Z, [Y, X] \rangle &= \\ = 2\langle Z, [X, Y] \rangle, & \end{aligned}$$

de donde se deduce que $\nabla_X Y - \nabla_Y X = [X, Y]$.

□

◇ **D.11.** Si consideramos \mathbb{R}^n con la métrica usual se cumple que la conexión Riemanniana queda dada por la derivada direccional, es decir,

$$(\nabla_X Y)_p = DY(p)[X_p]$$

D.4. Conexiones en subvariedades Riemannianas

Sea \tilde{M} variedad Riemanniana, $M \subset \tilde{M}$ una subvariedad y consideremos las conexiones de Levi-Civita asociadas ∇ y $\tilde{\nabla}$. Nos interesa saber como se relacionan, comenzaremos por observar que una no es “restricción” de la otra en el sentido siguiente:

Sea X_p en $T_p M$, Y en $\mathfrak{X}(M)$ y consideremos \tilde{Y} una extensión local de Y a \tilde{M} , sucede que si bien $(\tilde{\nabla}_X \tilde{Y})_p$ no depende de la extensión \tilde{Y} considerada (puesto que la dependencia local respecto de \tilde{Y} está dada por una derivada direccional en coordenadas locales), podría no encontrarse en $T_p M$ como en el ejemplo siguiente.

Ejemplo D.12. Sea $S^1 \subset \mathbb{R}^2$ con la métrica usual y $Y(x, y) = (y, -x)$.

Por lo cual $(\nabla_X Y)_p \neq (\tilde{\nabla}_X \tilde{Y})_p$.

Proposición D.13. $(\nabla_X Y)_p = P_{T_p M}((\tilde{\nabla}_X \tilde{Y})_p)$, siendo $P_{T_p M}$ la proyección ortogonal en $T_p M$.

Demostración. Veamos que la función $P_{T_p M}((\tilde{\nabla}_X \tilde{Y})_p)$ es una conexión y que verifica las propiedades de simetría y compatibilidad respecto de la métrica.

1.

$$\begin{aligned} (\nabla_{fX+gY} Z)_p &= P_{T_p M}(\tilde{\nabla}_{fX+gY} \tilde{Z}) = P_{T_p M}(f\tilde{\nabla}_X \tilde{Z} + g\tilde{\nabla}_Y \tilde{Z}) \\ &= fP_{T_p M}(\tilde{\nabla}_X \tilde{Z}) + gP_{T_p M}(\tilde{\nabla}_Y \tilde{Z}) = f(\nabla_X Z)_p + g(\nabla_Y Z)_p \end{aligned}$$

2. Análogo.

3.

$$\begin{aligned}
(\nabla_X fY)_p &= P_{T_p M}((\tilde{\nabla}_X f\tilde{Y})_p) = P_{T_p M}((\tilde{\nabla}_X f\tilde{Y})_p) \\
&= P_{T_p M}(X_p(f)\tilde{Y}_p + f(\tilde{\nabla}_X \tilde{Y})_p) \\
&= X_p(f)P_{T_p M}(\tilde{Y}_p) + fP_{T_p M}((\tilde{\nabla}_X \tilde{Y})_p) \\
&= X_p(f)P_{T_p M}(\tilde{Y}_p) + fP_{T_p M}((\tilde{\nabla}_X \tilde{Y})_p) = X_p(f)Y_p + f(\nabla_X Y)_p.
\end{aligned}$$

$$4. (\nabla_X Y - \nabla_Y X)_p = P_{T_p M}((\tilde{\nabla}_X \tilde{Y} - \tilde{\nabla}_Y \tilde{X})_p) = P_{T_p M}([\tilde{X}, \tilde{Y}]_p) = P_{T_p M}([\widetilde{X, Y}]_p) = [X, Y]_p$$

5. Utilizando que $\langle P_{T_p M}(\tilde{X}_p), Y_p \rangle = \langle \tilde{X}_p, Y_p \rangle$, obtenemos que

$$\begin{aligned}
(\langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle)_p &= \langle P_{T_p M}[(\tilde{\nabla}_{X_p} \tilde{Y})_p], Z_p \rangle + \langle Y_p, P_{T_p M}[(\tilde{\nabla}_{X_p} \tilde{Z})_p] \rangle \\
&= \langle (\tilde{\nabla}_{X_p} \tilde{Y})_p, Z_p \rangle + \langle Y_p, (\tilde{\nabla}_{X_p} \tilde{Z})_p \rangle \\
&= \langle (\tilde{\nabla}_{X_p} \tilde{Y})_p, \tilde{Z}_p \rangle + \langle \tilde{Y}_p, (\tilde{\nabla}_{X_p} \tilde{Z})_p \rangle \\
&= (\tilde{X} \langle \tilde{Y}, \tilde{Z} \rangle)_p = (X \langle Y, Z \rangle)_p.
\end{aligned}$$

□

Observación D.14. Sea $N \subset \mathbb{R}^n$ y consideremos \mathbb{R}^n con la métrica usual. Sabemos que $(\tilde{\nabla}_X \tilde{Y})_p = \frac{\partial \tilde{Y}}{\partial \tilde{X}_p}(p) = D\tilde{Y}(p)[X_p]$, luego,

$$(\nabla_X Y)_p = P_{T_p N}(D\tilde{Y}(p)[X_p])$$

D.4.1. Ejemplos

Ejemplo D.15. Consideremos \mathbb{R}^n con la métrica usual. Sea $S^{n-1} \subset \mathbb{R}^n$ con la métrica inducida. Tenemos que $P_{T_p S^{n-1}}(v) = (I_n - pp^t)v$, luego, de la observación anterior se desprende que

$$\nabla_{X_p} Y = P_{T_p S^{n-1}}(DY(p)[X_p]) = (I_n - pp^t)(DY(p)[X_p])$$

Ejemplo D.16. Sea $St(k, n) \subset \mathbb{R}^{n \times k} \cong \mathbb{R}^{nk}$ con la métrica inducida. Tenemos que $P_{T_N St(k, n)}(v) = (I_n - NN^t)v + Nskew(N^t v)$, luego, de la observación anterior se desprende que

$$\begin{aligned}\nabla_{X_N} Y &= P_{T_N St(p, n)}(DY(N)[X_N]) \\ &= (I_n - NN^t)DY(N)[X_N] + Nskew(N^t DY(N)[X_N])\end{aligned}$$

D.5. Conexión Riemanniana en espacios homogéneos

Supongamos que G un grupo de Lie con una estructura Riemanniana. Sea H un subgrupo de Lie cerrado. Supongamos que G/H es un espacio homogéneo para el cual puede inducirse la métrica Riemanniana (ver Sección C.6).

Si $X, Y \in \mathfrak{X}(G/H)$, denotaremos por $\hat{X}, \hat{Y} \in \mathfrak{X}(G)$ a los campos levantados sobre G que son horizontales.

Proposición D.17. Sean ∇ y $\tilde{\nabla}$ las conexiones Riemannianas sobre G y sobre G/H respectivamente. Entonces

$$\tilde{\nabla}_Y X = P^H(\nabla_{\hat{Y}} \hat{X})$$

siendo $X, Y \in \mathfrak{G}/\mathfrak{H}$, P^H la proyección orthogonal en el espacio horizontal.

Este resultado es de mucha utilidad para poder calcular conexiones sobre espacios homogéneos. Lo que dice es que el levantado horizontal de la derivada covariante de X en la dirección de Y coincide con la proyección orthogonal en el espacio horizontal de la derivada covariante del levantado de X en la dirección del levantado de Y .

◇ **D.18.** Dar fórmulas para la conexión de la variedad de Stiefel como espacio homogéneo.

D.6. Campo aceleración

Sea M una variedad Riemanniana y ∇ conexión afín. Sea γ curva en M con dominio I . Denotaremos $\mathfrak{X}(\gamma)$ al conjunto de campos en γ , es decir

$$\mathfrak{X}(\gamma) = \{X : \text{Im}(\gamma) \rightarrow TM : X(\gamma(t)) \in T_{\gamma(t)}M\}$$

Lema D.19. *Sea γ curva en M , y X, Y campos definidos en M . Denotemos por $p := \gamma(0)$, y $v := \dot{\gamma}(0)$. Si X, Y son campos vectoriales definidos en un entorno de γ tales que coinciden sobre γ , i.e. $X \circ \gamma = Y \circ \gamma$, entonces $\nabla_v X = \nabla_v Y$.*

Demostración. En coordenadas locales, con base local $\{D_1, \dots, D_n\}$, podemos escribir $X = \sum_i \alpha_i D_i$, y $Y = \sum_j \beta_j D_j$. En particular se tiene que $\alpha_i \circ \gamma = \beta_i \circ \gamma$. Luego, resulta

$$\begin{aligned} (\nabla_v X)_{\gamma(0)} &= \sum_i \nabla_v (\alpha_i D_i)_{\gamma(0)} \\ &= \sum_i \frac{d}{dt} (\alpha_i \circ \gamma)|_p \cdot D_i + \alpha_i (\nabla_v D_i)_{\gamma(0)} \\ &= \sum_i \frac{d}{dt} (\beta_i \circ \gamma)|_p \cdot D_i + \beta_i (\nabla_v D_i)_{\gamma(0)} = (\nabla_v Y)_{\gamma(0)} \end{aligned}$$

□

Este resultado muestra que tiene sentido definir la derivada de un campo x a lo largo de una curva. En particular, $\nabla_v X$ sólo depende del valor de X a lo largo de γ .

Dado que todo campo sobre una curva puede extenderse localmente (sin demostración), el Lema D.19 nos permite hacer la siguiente definición.

Definición D.20 (Derivada covariante). *Sea γ curva en M . Se define *derivada covariante* de un campo $X \in \mathfrak{X}(\gamma)$, y lo denotamos por $\frac{DX}{dt}$, al campo definido por $\nabla_{\dot{\gamma}(t)} \hat{X}$, siendo \hat{X} alguna extensión local de X en un entorno de $\gamma(t)$.*

Proposición D.21. *La función $\frac{D}{dt} : \mathfrak{X}(\gamma) \rightarrow \mathfrak{X}(\gamma)$ así definida verifica las siguientes propiedades*

1. \mathbb{R} -linealidad

$$\frac{D}{dt}(aX + bY) = a \frac{DX}{dt} + b \frac{DY}{dt}$$

2. *Regla de Leibniz*

$$\frac{D}{dt}(fX) = f'X + f\frac{DX}{dt}$$

3. Si $X \in \mathfrak{X}(M)$, entonces

$$\frac{D}{dt}(X \circ \gamma)(t) = (\nabla_{\dot{\gamma}(t)}X)_{\gamma(t)}$$

◇ **D.22.** Dar una prueba de este resultado.

Un caso muy especial, y por el cual se introduce la noción de conexión, es poder definir la noción de aceleración de una curva.

Definición D.23 (Aceleración de una curva). El *campo aceleración* de γ se define como

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

A veces se utiliza la notación $\nabla_{\dot{\gamma}(t)}\dot{\gamma}(t)$.

Observación D.24. Podemos obtener una expresión para $\frac{D^2}{dt^2}(\gamma)$ en coordenadas locales. Dada una carta $\phi : V_p \subset M \rightarrow U \subset \mathbb{R}^n$ y sea $\phi(\gamma(t)) = (x_1(t), \dots, x_n(t))$, sabemos que $\dot{\gamma}(t) = \sum_i \dot{x}_i(t)D_i(\gamma(t))$

$$\begin{aligned} \frac{D^2}{dt^2}(\gamma) &= \frac{D}{dt}(\dot{\gamma}) = \frac{D}{dt} \left(\sum_i \dot{x}_i(t)D_i(\gamma(t)) \right) \\ &= \sum_{i=1}^n \ddot{x}_i(t)D_i(\gamma(t)) + \sum_{i,j=1}^n \dot{x}_i(t)\dot{x}_j(t)(\nabla_{D_j}D_i)(\gamma(t)). \end{aligned} \quad (\text{D.3})$$

◇ **D.25.** De (D.3) resulta que γ es geodésica si y sólo si las coordenadas locales satisfacen la ecuación diferencial de segundo orden

$$\ddot{x}_k + \sum_{i,j} \Gamma_{i,j}^{(k)} \dot{x}_i \dot{x}_j = 0, \quad (k = 1, \dots, m).$$

Luego por teorema de existencia y unicidad de soluciones de ecuaciones diferenciales resulta que existen geodésicas definidas localmente con condiciones iniciales $p = \gamma(0)$ y $v = \dot{\gamma}(0)$.

Definición D.26 (Geodésica). Una curva γ en M se dice *geodésica* si $\frac{D^2}{dt^2}(\gamma) = \mathbf{o}$ para todo t .

◇ **D.27.** sea $M \subset \mathbb{R}^n$ una variedad encajada con la métrica Riemanniana inducida por el ambiente. Sea γ una geodésica en M . Entonces el campo aceleración de γ , como curva euclídea, es ortogonal a M . (cf. Sección C.4.)

◇ **D.28.** Extender el resultado anterior a M subvariedad Riemanniana de una variedad Riemanniana \widehat{M} .

◇ **D.29.** Sea M una variedad Riemanniana con ∇ la conexión compatible. Sea γ curva en M , y sean $X, Y \in \mathfrak{X}(\gamma)$. Probar que

$$\frac{d}{dt}\langle X, Y \rangle = \left\langle \frac{DX}{dt}, Y \right\rangle + \left\langle X, \frac{DY}{dt} \right\rangle \quad (\text{D.4})$$

Observación D.30. Para todo p en M y X_p en T_pM existe una única curva γ geodésica tal que $\gamma(0) = p$ y $\dot{\gamma}(0) = X_p$. La notación utilizada será $\gamma(t) = \gamma(t; p, X_p)$.

Definición D.31 (Mapa exponencial). Denominaremos *mapa exponencial* al mapa $Exp_p : T_pM \rightarrow M$ definido como

$$Exp_p(x_p) = \gamma(1; p, x_p).$$

Comentario D.32. No siempre sucede que las geodésicas están definidas para todo tiempo, y por lo tanto el mapa Exp_p no tiene por qué estar definido sobre todo el espacio tangente T_pM . (¿Podrías dar un ejemplo?) Cuando ocurre que el dominio de Exp_p es todo T_pM , para todo $p \in M$, se dice que M es una variedad Riemanniana *completa*.

Utilizaremos el siguiente resultado sin demostración.

Teorema D.33. *El mapa exponencial es una retracción.* □

Proposición D.34. *La curva $t \mapsto Exp_p(tX_p)$, es la geodésica que pasa por p con velocidad X_p .*

Demostración. Dado que Exp_p es una retracción se tiene que

$$\begin{aligned} \frac{d}{dt}Exp_p(tX_p)\Big|_{t=0} &= DExp_p(tX_p) \cdot X_p\Big|_{t=0} \\ &= DExp_p(\mathbf{o}) \cdot X_p = X_p \end{aligned}$$

□

D.7. Hessiano

Sea M una variedad Riemanniana y ∇ la conexión Riemanniana.

Definición D.35. El *Hessiano* de la función f en $C^\infty(M)$ en el punto p es una transformación lineal $\text{Hess}_p f : T_p M \rightarrow T_p M$ que verifica

$$\text{Hess}_p f[X_p] = \nabla_{X_p} \text{grad}(f).$$

El siguiente ejercicio muestra que con esta definición coincide con el Hessiano euclideo.

◇ **D.36.** Sea $M = \mathbb{R}^n$ con conexión Riemanniana ∇ , y $f \in C^\infty(M)$. Probar que

$$\langle \text{Hess}_p f[X_p](u), v \rangle = D^2 f(p)(u, v), \quad (\text{D.5})$$

siendo $D^2 f$ el operador derivada segunda de f . Observar que en particular, si $\{e_1, \dots, e_n\}$ es la base canónica, entonces resulta

$$\langle \text{Hess}_p f[X_p](e_i), e_j \rangle = \frac{\partial^2 f}{\partial x_i \partial x_j}(p) \quad (i, j = 1, \dots, n).$$

Veamos que el Hessiano así definido verifica ciertas propiedades que el Hessiano usual también verifica y que son de nuestro interés particular.

Lema D.37. El Hessiano verifica la siguiente fórmula para todo X, Y en $\mathfrak{X}(M)$.

$$\langle \text{Hess} f[X], Y \rangle = X(Y(f)) - (\nabla_X Y)f.$$

Demostración. $\langle \text{Hess} f[X], Y \rangle = \langle \nabla_X \text{grad}(f), Y \rangle$, por ser una conexión compatible con la métrica tenemos que $\langle \nabla_X \text{grad}(f), Y \rangle = X \langle \text{grad}(f), Y \rangle - \langle \text{grad}(f), \nabla_X Y \rangle = X(Y(f)) - (\nabla_X Y)(f)$ □

Proposición D.38. El Hessiano es simétrico, es decir, para todo X, Y en $\mathfrak{X}(M)$ se cumple que

$$\langle \text{Hess} f[X], Y \rangle = \langle X, \text{Hess} f[Y] \rangle.$$

Demostración. Por el lema anterior tenemos que $\langle \text{Hess } f[X], Y \rangle = X(Y(f)) - (\nabla_X Y)(f)$ y $\langle \text{Hess } f[Y], X \rangle = Y(X(f)) - (\nabla_Y X)(f)$. A su vez, por la simetría de la conexión se tiene que

$$\begin{aligned} \langle \text{Hess } f[X], Y \rangle &= X(Y(f)) - (\nabla_X Y)(f) = X(Y(f)) - ((\nabla_Y X)(f) + [X, Y](f)) = \\ &= X(Y(f)) - [(\nabla_Y X)(f) + X(Y(f)) - Y(X(f))] = \langle \text{Hess } f[Y], X \rangle. \end{aligned}$$

□

Observación D.39. La función $B: T_p M \times T_p M \rightarrow \mathbb{R}$ dada por $B(X_p, Y_p) = \langle \text{Hess}_p f[X_p], Y_p \rangle$ es una forma bilineal simétrica

Proposición D.40. *Se tiene*

$$\langle \text{Hess } f_p[u], u \rangle = \frac{d^2}{dt^2} (f \circ \text{Exp}_p(tX_p)) \Big|_{t=0}$$

Demostración. Observar que

$$\begin{aligned} \frac{d^2}{dt^2} (f \circ \text{Exp}_p)(tX_p) \Big|_{t=0} &= \frac{d}{dt} \left[\frac{d}{dt} [f(\text{Exp}_p(tX_p))] \right] \Big|_{t=0} \\ &= \frac{d}{dt} \left[Df(\text{Exp}_p(tX_p)) \left(\frac{d}{dt} \text{Exp}_p(tX_p) \right) \right] \Big|_{t=0} \\ &= \frac{d}{dt} \left\langle \text{grad}(f(\text{Exp}_p(tX_p))), \frac{d}{dt} \text{Exp}_p(tX_p) \right\rangle \Big|_{t=0}. \end{aligned}$$

De (D.4) se tiene

$$\begin{aligned} &\frac{d}{dt} \left\langle \text{grad}(f(\text{Exp}_p(tX_p))), \frac{d}{dt} \text{Exp}_p(tX_p) \right\rangle \Big|_{t=0} = \\ &= \left(\left\langle \frac{D}{dt} \text{grad}(f(\text{Exp}_p(tx_p))), \frac{d}{dt} \text{Exp}_p(tx_p) \right\rangle + \left\langle \text{grad}(f(\text{Exp}_p(tx_p))), \frac{D^2}{dt^2} \text{Exp}_p(tx_p) \right\rangle \right) \Big|_{t=0}. \end{aligned}$$

Utilizando la Proposición D.34, y que $\text{Exp}_p(tx_p)$ es una geodésica se concluye

$$\frac{d^2}{dt^2} (f \circ \text{Exp}_p)(tX_p) \Big|_{t=0} = \langle \nabla_{X_p} \text{grad}(f), X_p \rangle,$$

lo que termina la prueba. □

En algunos textos se utiliza el resultado de la Proposición D.40 como definición del Hessiano (ver por ejemplo [EAS]).

Corolario D.41. $\text{Hess}_p f = \text{Hess}_o(f \circ \text{Exp}_p)$

Demostración. Probaremos que $\langle \text{Hess}_p f[X_p], Y_p \rangle = \langle \text{Hess}_o(f \circ \text{Exp}_p)[X_p], Y_p \rangle$ para todo X_p, Y_p . Basta probarlo para $X_p = Y_p$ dado que $B(X_p, Y_p)$ definida como en la observación es una forma bilineal simétrica. Por propiedades del Hessiano usual se tiene que

$$\langle \text{Hess}_o(f \circ \text{Exp}_p)[X_p], X_p \rangle = \frac{d^2}{dt^2}(f \circ \text{Exp}_p)(tX_p) \Big|_{t=0}.$$

Luego el resultado sigue de la Proposición D.40. □

Bibliografía

- [AMT] P.-A. Absil, Robert Mahony, and Jochen Trumpf, *An extrinsic look at the Riemannian Hessian*, Geometric science of information.
- [AMS] Absil, Mahony, and Sepulchre, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press.
- [A] Gregory S. Ammar, *Geometric aspects of Hessianberg matrices*, Differential geometry: the interface between pure and applied mathematics (San Antonio, Tex., 1986).
- [AM] Gregory Ammar and Clyde Martin, *The geometry of matrix eigenvalue methods*, Acta Appl. Math. **5** (1986), no. 3, 239–278. MR868890
- [B] Andrew Baker, *Matrix groups*, Springer Undergraduate Mathematics Series, Springer-Verlag London, Ltd., London, 2002. An introduction to Lie group theory. MR1869885
- [BCSS] Lenore Blum, Felipe Cucker, Michael Shub, and Steve Smale, *Complexity and real computation*, Springer-Verlag, New York, 1998. With a foreword by Richard M. Karp. MR1479636
- [B] William M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Second, Pure and Applied Mathematics, vol. 120, Academic Press, Inc., Orlando, FL, 1986.
- [C] Ralph Cohen, *Topics in Morse theory: lecture notes*. Notas de clase.
- [D] Jean-Pierre Dedieu, *Points fixes, zéros et la méthode de Newton*, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 54, Springer, Berlin, 2006 (French). With a preface by Steve Smale.
- [DNT] P. Deift, T. Nanda, and C. Tomei, *Ordinary differential equations and the symmetric eigenvalue problem*, SIAM J. Numer. Anal. **20** (1983), no. 1, 1–22.
- [D] Shaoqiang Deng, *Homogeneous Finsler spaces*, Springer Monographs in Mathematics, Springer, New York, 2012. MR2962626

- [dC] Manfredo Perdigão do Carmo, *Riemannian geometry*, Mathematics: Theory & Applications, Birkhäuser Boston, Inc., Boston, MA, 1992. Translated from the second Portuguese edition by Francis Flaherty.
- [DKV] J. J. Duistermaat, J. A. C. Kolk, and V. S. Varadarajan, *Functions, flows and oscillatory integrals on flag manifolds and conjugacy classes in real semisimple Lie groups*, *Compositio Math.* **49** (1983), no. 3, 309–398. MR707179
- [EAS] Alan Edelman, Tomás A. Arias, and Steven T. Smith, *The geometry of algorithms with orthogonality constraints*, *SIAM J. Matrix Anal. Appl.* **20** (1999), no. 2, 303–353. MR1646856
- [H1] Mads Hansen, *Morse Theory on Complex Grassmannians*, Construction from the classical outset to Morse-Bott theory. Notas de clase.
- [H2] Ralph Howard, *Analysis on Homogeneous Spaces*, Royal Institute of Technology Stockholm. Notas de clase.
- [H3] ———, *The kinematic formula in Riemannian homogeneous spaces*, *Mem. Amer. Math. Soc.* **106** (1993), no. 509, vi+69. MR1169230
- [GHL] Sylvestre Gallot, Dominique Hulin, and Jacques Lafontaine, *Riemannian geometry*, 3rd ed., Universitext, Springer-Verlag, Berlin, 2004.
- [IMKNZ] Arieh Iserles, Hans Z. Munthe-Kaas, Syvert P. Nørsett, and Antonella Zanna, *Lie-group methods*, *Acta numerica*, 2000.
- [K1] Narendra Karmarkar, *Riemannian geometry underlying interior-point methods for linear programming*, *Mathematical developments arising from linear programming* (Brunswick, ME, 1988).
- [K2] Takashi Koda, *An introduction to the geometry of homogeneous spaces*, *Proceedings of the 13th International Workshop on Differential Geometry and Related Fields [Vol. 13]*.
- [HM] Helmke and Moore, *Optimization and Dynamical Systems*, Springer.
- [P] Peter Petersen, *Riemannian geometry*, 3rd ed., *Graduate Texts in Mathematics*, vol. 171, Springer, Cham, 2016. MR3469435
- [T] Carlos Tomei, *The Toda lattice, old and new*, *J. Geom. Mech.* **5** (2013), no. 4, 511–530. MR3180710
- [S] Smith, *Optimization Techniques on Riemannian Manifolds*, *Fields Institute Communication*.
- [Z] Michael M and Pappas Zavlanos George J, *A dynamical systems approach to weighted graph matching*, *Automatica*, 44, 11, 2817–2824, Elsevier, 2008.

Índice alfabético

- Álgebra de Lie, [120](#), [135](#)
- Acción de grupos, [138](#)
 - órbita, [139](#)
- Algoritmo
 - QR, [35](#)
- Algoritmos
 - Cociente de Rayleigh, [98](#)
 - Método de Newton para campos vectoriales, [96](#)
 - Minimización Grass(p, n), [99](#)
 - Newton para puntos críticos, [97](#)
 - QR, [36](#)
- Algoritmos: Armijo modificado, [79](#)
- Algoritmos: búsqueda lineal, [78](#)
- Aproximación de Euler, [94](#)
- Campo de vectores, [118](#)
 - completo, [121](#)
 - flujo local, [121](#)
- Campo vectorial
 - gradiente, [128](#)
- Cartas compatibles, [114](#)
- Clase de equivalencia, [104](#)
- Cociente de Rayleigh, [26](#)
 - algoritmo, [98](#)
 - generalizado, [30](#)
- Conexiones
 - definición, [148](#)
 - derivada covariante, [156](#)
 - euclideana, [148](#), [150](#)
 - Riemanniana, [151](#)
 - geodésicas, [158](#)
 - símbolos de Christoffel, [149](#)
 - simétrica, [151](#)
- Conjugación topológica, [17](#)
- Coordenadas locales, [113](#)
- Corchete de Lie, [120](#)
 - campo vectorial, [125](#)
 - matrices, [38](#)
- Curva
 - aceleración, [157](#)
 - integral, [120](#)
- Derivación, [119](#)
- Descomposición
 - de Schur, [25](#)
 - en valores singulares, [40](#)
 - QR, [35](#)
- Ecuación de Brockett, [58](#)
 - esfera, [59](#)

- Ecuación de Riccati
 una variable, 20
 vectorial, 19
- Espacio bandera, 23
 acción de $GL(n, \mathbb{C})$, 24
- Espacio cociente, 104
 propiedad universal, 111
 proyección cociente, 104
 Relación de equivalencia, 103
 topología cociente, 111
- Espacio homogéneo, 140
- Espacio horizontal
 distribución, 77
- Espacio métrico
 bola, 109
 métrica, 107
 métrica inducida, 108
 topología métrica, 109
- Espacio normado, 108
- Espacio proyectivo, 17
- Espacio topológico, 104
 compacto, 110
 cubrimiento, 109
 Hausdorff, 105
 producto, 107
 fibras, 107
 inclusión canónica, 107
 proyecciones, 107
 propiedad Bolzano-Weierstrass, 110
 secuencialmente compacto, 110
- Fibrado tangente, 118
 proyección canónica, 118
- Flujo
 gradiente, 129
 Rayleigh, 27
- isoespectral, 38
 isosingular, 41
 Toda, 51
- Función
 continua, 105
 diferenciable, 114
 valor crítico, 122
 valor regular, 122
 encaje, 123
 inmersión, 123
 Morse-Bott, 126
 submersión, 123
- Gradiente, 128
- Grafos, 65
 adyacencia, 67
 automorfismo, 66
 isomorfos, 65
- Grassmaniana
 espacio horizontal, 88
 espacio vertical, 88
 geodésica, 89
 gradiente, 90
 métrica, 88
- Grupo
 representación, 139
 a un parámetro, 121
 acción, 138
- Grupo de Lie, 133
 mapa exponencial, 137
 campo invariante, 135
 grupo lineal especial, 134
 grupo ortogonal, 134
 grupo ortogonal especial, 135
 grupo unitario, 135
 grupo unitario especial, 135

- subgrupo de Lie, 134
 - traslación a izquierda, 135
- Grupos
 - isometrías lineales, 135
- Hessiano, 124
- Integral de un sistema Hamiltoniano, 52
- Inversa de Moore Penrose, 42
- Lema de Morse, 126
- Método de la potencia, 16, 17
 - en $\text{Flag}(\mathbb{C}^n)$, 23
 - en $\mathbb{P}(\mathbb{C}^n)$, 17
- Método de Newton, 93
 - operador, 94
- Métrica Riemanniana, 127
- Métricas bi-invariantes, 144
- Matrices
 - Hessenberg, 50
 - logaritmo, 20
 - mapa exponencial, 19
 - parte antisimétrica, 32
 - parte simétrica, 32
 - producto interno de Frobenius, 30
- Mínimos cuadrados, 41
- Morse-Bott, 126
- Problemas
 - automorfismos de grafos, 66
 - Graph Matching Problem, 67
 - isomorfismo de grafos, 66
 - programación lineal, 57
 - subespacios invariantes, 21
 - vector propio dominante, 15
- Programación lineal, 57
 - polítopo, 62
- Simplex, 59
- Pseudoinversa, 42
- Punto crítico, 123
 - subvariedad crítico, 123
- Regla de Leibniz, 115
- Representación de un grupo, 139
- Retracción, 73
- Retracciones
 - S^{n-1} , 75
 - $St(k, n)$, 76
 - $\text{Grass}(k, n)$, 78
 - $\mathcal{O}(n)$, 75
 - $\mathbb{P}(\mathbb{R}^n)$, 77
- Símbolos de Christoffel, 149
- Simplex, 59
- Subespacio dominante, 29
- Sucesión de Cauchy, 109
- Sucesión gradiente relacionado, 78
- Teorema
 - Cartan, 134
 - Courant-Fisher, 26
 - de las sumersiones, 122
 - dinámica flujo gradiente, 130
 - dinámica QR, 37
 - flujo corchete doble, 44
 - Flujo de Brockett en S^{n-1} , 60
 - flujo generalizado de Rayleigh, 34
 - flujo gradiente de Rayleigh generalizado, 35
 - flujo QR, 39
 - flujo Toda, 53
 - Shub-Vasquez, 25
- Topología, 104
 - base, 106

borde, frontera, 105
cerrado, 104
comparación, 105
continuidad en un punto, 105
entorno, 104
punto de acumulación, 104
punto interior, 105
sucesión convergente, 105
topología final, 111

subvariedad, 122
 inmersa, 123
vector tangente, 115
Variedad Riemanniana, 127
 completa, 158
 mapa exponencial, 158
Vector propio dominante, 15

Variedad

 ortogonalmente semejantes, 43, 54

Variedad de Stiefel

 definición, 29, 30
 espacio horizontal, 86
 espacio normal, 32
 espacio vertical, 86
 estructura diferenciable, 30
 métrica euclídea, 82
 métrica homogénea, 87
 métrica Riemanniana, 32
 tangente, 31

Variedad diferenciable, 114

 atlas diferenciable, 114
 cambio de coordenadas, 114
 campo de vectores, 118
 Carta local, 113
 cociente
 espacio horizontal, 77
 curva, 116
 difeomorfismo, 114
 diferencial de mapas, 116
 espacio tangente, 115
 fibrado tangente, 118
 función diferenciable, 114
 mapa de transición, 113