



CIMAT

Centro de Investigación en Matemáticas, A.C.

---

# DE LA MECÁNICA GEOMÉTRICA A LOS ALGORITMOS DE OPTIMIZACIÓN GEOMÉTRICA

**T E S I S**

Que para obtener el grado de  
**Maestro en Ciencias**  
con Orientación en  
**Matemáticas Básicas**

Presenta

Jeisson Andrés Molano Guzmán

Director de Tesis:

Dr. Alessandro Bravetti

*Alessandro Bravetti*

---

Autorización de la versión final



## AGRADECIMIENTOS

Agradezco a Alessandro Bravetti por aceptar ser mi asesor de tesis, por toda su paciencia y cantidad de horas invertidas ayudándome a entender los temas y revisando el presente trabajo, que a pesar que gran parte del trabajo fue a distancia siempre fue paciente, entusiasta y motivador.

También agradezco a Raúl Quiroga Barranco (asesor de tesis) y Jimmy Petean por ser parte del comité evaluador y por sus correcciones hechas al presente documento.

A Juan Pablo Saya, Paola Bastidas, James Cortes antiguos estudiantes míos de Tumaco, a mis padres y a Felipe Rodríguez, una persona muy especial, que siempre estuvieron apoyándome, escuchándome y dándome ánimos para continuar.

A Alejandro Pérez, Dulce Silva, Issis Fragoso compañeros de CIMAT con los que compartí demasiado durante toda mi estancia en CIMAT e hicieron muy agradable mi tiempo aquí.

A CIMAT que me dio las mejores condiciones y facilidades para realizar mis estudios de maestría con agrado.

Finalmente agradezco al Consejo Nacional de Ciencia y Tecnología (CONACYT) por otorgarme la beca de postgrado (No. 940846) que hizo posible que este trabajo se pudiera realizar.

<b>Introducción</b>		<b>3</b>
<b>1. Mecánica Geométrica</b>		<b>4</b>
1.1. Una Introducción a la Geometría Simpléctica y de Contacto		4
1.1.1. Geometría Simpléctica		4
1.1.2. Geometría de Contacto		8
1.2. Ecuaciones de Hamilton para Sistemas Conservativos		10
1.3. Ecuaciones de Hamilton para Sistemas Disipativos		15
1.3.1. Hamiltonianos dependientes del tiempo		16
1.3.2. Hamiltonianos conformalmente simplécticos		18
1.3.3. Hamiltonianos de contacto		19
<b>2. Geometría de la Información</b>		<b>23</b>
2.1. Variedades Estadísticas		23
2.2. Funciones de Divergencia		25
2.3. Información de Fisher		29
<b>3. Algoritmos de Optimización Geométrica</b>		<b>35</b>
3.1. Introducción a la Optimización		35
3.1.1. Métodos con el gradiente		35
3.1.2. Métodos con el gradiente más momentum		36
3.1.3. De los algoritmos de optimización a los sistemas dinámicos		37
3.2. Optimización Hamiltoniana		40
3.2.1. Hamiltonianos dependientes del tiempo		40
3.2.2. Hamiltonianos conformalmente simplécticos		42
3.2.3. Hamiltonianos de contacto		44
3.3. Optimización Hamiltoniana Natural		45
<b>Conclusiones</b>		<b>54</b>
<b>Bibliografía</b>		<b>56</b>
<b>A. Integración Geométrica</b>		<b>58</b>

La elección del tema de este trabajo se dio cuando mi asesor de tesis Alessandro me mostró el artículo de [LZ17]. Al ver el nombre del artículo me llamó la atención el nombre de “geometría de la información” (de la cual nunca había escuchado) y además que tenía una conexión con la mecánica. Sabía que la mecánica Hamiltoniana traía una estructura simpléctica la cual nunca estudié en mis cursos de licenciatura, solo había asistido a algunas conferencias que hablaban sobre el tema, pero nunca lo estudié a fondo. Además, debido a una charla dada por mi asesor sobre geometría de contacto aplicada a la termodinámica en CIMAT, hizo que me interesara en estudiar estas estructuras geométricas que tienen aplicaciones en la física. Así que aproveché este momento para estudiar las estructuras simplécticas y de contacto las cuales me permitieron adentrarme en la geometría de la información. El texto de [Mc15], el cual encontré mientras paseaba en la biblioteca de CIMAT, me ayudó bastante a realizar este objetivo ya que tiene muchos ejemplos y una buena exposición básica en geometría Riemanniana, simpléctica y de contacto, así que fue un excelente inicio para mí, que no conocía muchos de estos temas. Este texto hizo que fuera más fácil entender el texto de [dS01] que es uno de los textos clásicos que se suele utilizar para abordar los temas de geometría simpléctica y de contacto.

El objetivo general del trabajo fue estudiar la relación entre la geometría de la mecánica y la geometría de la información, para ello se utilizó las funciones de divergencia, las cuales permiten inducir una estructura simpléctica y de contacto sobre las variedades en uso en geometría de la información. Este es el primer paso para construir una dinámica Hamiltoniana sobre dichas variedades, lo cual hemos aplicado luego para construir algoritmos de optimización de forma geométrica.

En este trabajo realizamos la construcción de algoritmos de optimización geométricos. Esta línea de investigación ha crecido mucho en los últimos años, en especial en la comunidad que trabaja con *Machine Learning*, donde los algoritmos de optimización son fundamentales para entrenar la máquina. Se ha visto en particular que muchos de los algoritmos más usados, como por ejemplo el momento clásico o el método del gradiente acelerado de Nesterov son discretizaciones de sistemas dinámicos Hamiltonianos con diferentes tipos de disipación como se puede ver en [SBC16]. Esto ha motivado empezar la investigación desde la perspectiva geométrica y de los sistemas dinámicos, lo cual facilita el estudio de las tasas de convergencia, y luego usar los métodos de discretización geométricos para obtener algoritmos que estén garantizados preservar dichas tasas por ejemplo ver [EJV20].

Por estas razones en este trabajo nos enfocamos en la parte geométrica y en probar la convergencia a la solución óptima en el caso continuo, y sólo mencionamos la estructura general de los métodos de discretización que luego se pueden usar para transformar estos sistemas en algoritmos que se puedan implementar.

Los resultados principales del trabajo son el Teorema [3.3.4](#), el cual es una extensión natural del Teorema 1 en [\[LZ17\]](#) al caso de contacto, y el Teorema [3.3.8](#), en el cual se propone un sistema Hamiltoniano de contacto que se construye a partir de las divergencias de Bregman y se prueba que el flujo de este sistema converge al mínimo de la función convexa que se pretende optimizar con una tasa de convergencia exponencial, extendiendo y aclarando algunos de los resultados de [\[WWJ16\]](#).

La estructura de la tesina es la siguiente:

En el capítulo 1 estudiamos las estructuras simplécticas y de contacto y nos ayudamos de los textos de [\[Mc15\]](#) y [\[dS01\]](#) para escribir esta parte. Utilizaremos estas estructuras para construir la mecánica Hamiltoniana y consideraremos los casos conservativos y disipativos, donde consideraremos tres formas distintas de hacer disipación de energía. Toda esta parte es tomada de [\[BCT17\]](#), [\[McP01\]](#) y [\[FSRV19\]](#), estos dos últimos para la parte de Hamiltonianos conformalmente simplécticos.

El capítulo 2 lo dedicamos a estudiar la geometría de la información, donde definimos una variedad estadística y estudiamos su estructura y propiedades. Además, introducimos las funciones de divergencia en donde el Teorema [2.2.2](#) es el más importante para nuestro documento, pues permite construir métricas Riemannianas sobre variedades diferenciables, como por ejemplo las geometrías inducidas por la divergencia de Bregman, divergencia canónica y la divergencia de Kullback-Leibler la cual nos genera una geometría que tiene una métrica generada por la matriz de información de Fisher. El contenido de este capítulo es tomado de [\[A16\]](#), [\[L19\]](#) y [\[N18\]](#). Cabe mencionar que estas estructuras tienen nociones de longitud, curvatura, geodésicas y demás términos conocidos en geometría Riemanniana que no mencionaremos debido a que no los utilizaremos para el objetivo de este documento. El lector interesado puede consultar esto en [\[A16\]](#) y [\[N18\]](#).

En el capítulo 3 nos dedicamos a ver métodos de optimización. Comenzamos introduciendo los métodos del gradiente, gradiente natural, momento clásico y Nesterov de los cuales encontramos sus límites continuos utilizando la idea de [\[SBC16\]](#). Estos límites continuos dan ecuaciones de sistemas mecánicos disipativos y por tanto es natural pensar en hacer con ellos optimización y de ahí tenemos la sección de optimización Hamiltoniana donde usamos lo visto en el capítulo 1 y seguimos las ideas de [\[BJW18\]](#), [\[BDFB19\]](#), [\[BSVZ19\]](#), [\[FSRV19\]](#), [\[EJV20\]](#) y [\[WWJ16\]](#). Para finalizar tomamos el teorema principal de [\[LZ17\]](#), Teorema [3.3.1](#), el cual induce una 2-forma simpléctica sobre la variedad  $M \times M$  con la ayuda de una función de divergencia, con esto encontramos un análogo que induce una 1-forma de contacto sobre  $M \times M \times \mathbb{R}$  y que utilizamos para hacer optimización continua. Además, aquí tenemos los Teoremas [3.3.4](#) y [3.3.8](#) que como mencionamos antes son los resultados principales del trabajo.

Finalmente se agrega un apéndice en el cual se discuten sobre integración numérica, ha-

ciendo énfasis sobre los métodos de discretización simplécticos y de contacto. Además mencionamos el teorema clave de [E.IV20](#) que como se dijo antes da sentido que trabajemos con el caso continuo y no el discreto.

## 1.1. Una Introducción a la Geometría Simpléctica y de Contacto

### 1.1.1. Geometría Simpléctica

Las ideas claves de la geometría simpléctica comenzaron con los trabajos de Lagrange sobre mecánica celeste en 1808, pero fue posteriormente con los trabajos de Hamilton que se formalizaron estas ideas. Sin embargo, la formalización geométrica se realizó con los trabajos de Lie, Poincaré y Cartan. Así la geometría simpléctica ha sido considerada como la estructura asociada a la mecánica. Comenzaremos con la definición de una forma simpléctica utilizando las ideas de [dS01] y [Mc15].

**Definición 1.1.1.** Sea  $M$  una variedad. Definimos  $\omega$  una forma simpléctica sobre  $M$ , si  $\omega$  es una 2-forma (forma bilineal antisimétrica sobre  $T_x M$  para todo  $x \in M$ ) que cumple:

1.  $\omega$  es cerrada, es decir,  $d\omega = 0$ .
2.  $\omega$  es no degenerada, es decir, si para todo  $x \in M$  con  $u_x \in T_x M$  se tiene que  $\omega_x(u_x, v_x) = 0$  para todo  $v_x \in T_x M$  entonces  $u_x = 0$ .

La pareja  $(M, \omega)$  se llama **variedad simpléctica**. Note que las variedades simplécticas siempre deben tener dimensión par, ya que cualquier forma bilineal  $\omega$  en un espacio vectorial  $V$  de dimensión  $n$  puede ser escrita como  $\omega(u, v) = v^T A u$ , donde  $A$  es una matriz de tamaño  $n \times n$  y sus entradas dependen de la base elegida para  $V$ . Como la forma es antisimétrica así lo es  $A$  y solo en dimensión par se puede tener matrices antisimétricas y no degeneradas.

La siguiente proposición nos permite ver de una manera fácil cuando una 2-forma es no degenerada.

**Proposición 1.1.2.** Sea  $\omega$  una 2-forma en  $M$ .  $\omega$  es no degenerada si y solo si

$$\omega^n = \omega \wedge \cdots \wedge \omega \neq 0.$$

Ahora veremos unos ejemplos interesantes.



## 1.1. UNA INTRODUCCIÓN A LA GEOMETRÍA SIMPLÉCTICA Y DE CONTACTO 5

**Ejemplo 1.1.3.** Sea  $M = \mathbb{R}^{2n}$  y considere una carta coordenada  $(U, q_1, \dots, q_n, p_1, \dots, p_n)$ . Entonces la forma

$$\omega_0 = \sum_{i=1}^n dp_i \wedge dq_i \quad (1.1)$$

es simpléctica. En efecto, note que si definimos  $\alpha_0 = \sum_{i=1}^n p_i dq_i$ , se tiene que  $\omega_0 = d\alpha_0$  y así  $\omega_0$  es cerrada. Por otro lado  $\omega_0^n = n! dp_1 \wedge dq_1 \wedge \dots \wedge dp_n \wedge dq_n$  y por tanto al evaluar en la base  $\left\{ \frac{\partial}{\partial q_1}, \dots, \frac{\partial}{\partial q_n}, \frac{\partial}{\partial p_1}, \dots, \frac{\partial}{\partial p_n} \right\}$  de  $T_x U$ , se tiene como resultado  $n!$ , así  $\omega_0^n \neq 0$  sobre  $U$ .  $\omega_0$  se llama **forma simpléctica estándar**.

**Ejemplo 1.1.4.** Sean  $(M_1, \omega_1)$  y  $(M_2, \omega_2)$  variedades simplécticas con  $\dim(M_1) = 2m$  y  $\dim(M_2) = 2n$ . Sobre  $M_1 \times M_2$  vamos a definir una forma simpléctica dada por

$$\Omega = \pi_1^* \omega_1 + \pi_2^* \omega_2, \quad (1.2)$$

donde  $\pi_1 : M_1 \times M_2 \rightarrow M_1$  y  $\pi_2 : M_1 \times M_2 \rightarrow M_2$  son las funciones proyección. Veamos que  $\Omega$  es forma simpléctica.  $\Omega$  es una 2-forma desde la definición. Note que,  $d\Omega = d(\pi_1^* \omega_1 + \pi_2^* \omega_2) = \pi_1^* d\omega_1 + \pi_2^* d\omega_2 = 0$  por tanto  $\Omega$  es cerrada. Por otro lado

$$\begin{aligned} \Omega^{n+m} &= (\pi_1^* \omega_1 + \pi_2^* \omega_2)^{n+m} \\ &= \sum_{k=1}^{n+m} \binom{n+m}{k} (\pi_1^* \omega_1)^{m+n-k} \wedge (\pi_2^* \omega_2)^k \\ &= \binom{n+m}{n} (\pi_1^* \omega_1)^m \wedge (\pi_2^* \omega_2)^n, \end{aligned}$$

donde en la última igualdad se uso que  $(\pi_1^* \omega_1)^{m+n-k} = 0$  para  $m+n-k > m$  y  $(\pi_2^* \omega_2)^k = 0$  para  $k > n$ . Al escoger una carta coordenada se tiene que  $\Omega^{n+m} \neq 0$  en dicha carta, así  $\Omega$  es no degenerado.  $\Omega$  es llamada una **forma producto**.

**Ejemplo 1.1.5.** Siguiendo el ejemplo anterior si se define  $\Omega = \lambda_1 \pi_1^* \omega_1 + \lambda_2 \pi_2^* \omega_2$  donde  $\lambda_1, \lambda_2 \in \mathbb{R} - \{0\}$ , se tiene una forma simpléctica sobre el producto cartesiano de variedades. Tomando  $\lambda_1 = 1$  y  $\lambda_2 = -1$  se tiene

$$\tilde{\Omega} = \pi_1^* \omega_1 - \pi_2^* \omega_2, \quad (1.3)$$

la cual llamamos **forma producto torcido**.

**Definición 1.1.6.** Sean  $(M_1, \omega_1)$  y  $(M_2, \omega_2)$  variedades simplécticas  $2n$ -dimensionales, y  $g : M_1 \rightarrow M_2$  un difeomorfismo. Si  $g^* \omega_2 = \omega_1$  entonces  $g$  se llama **simplectomorfismo**.

El siguiente teorema me dice que toda variedad simpléctica es localmente simplectomorfa a  $(\mathbb{R}^{2n}, \omega_0)$ . Esto es muy importante pues siempre se puede trabajar localmente con un espacio que conocemos bien.

**Teorema 1.1.7.** (Darboux) Sea  $(M, \omega)$  una variedad simpléctica  $2n$ -dimensional, y sea  $x \in M$ . Entonces existe una carta coordenada  $(U, q_1, \dots, q_n, p_1, \dots, p_n)$ ,  $x \in U$  y un difeomorfismo  $\phi : U \rightarrow \mathbb{R}^{2n}$  tal que  $\phi^* \omega_0 = \omega$  en  $U$ , donde  $\omega_0$  esta dado por **(1.1)**.

Algo interesante es que el fibrado cotangente de toda variedad tiene una 1-forma que aparece de manera natural y la cual va a permitir definir una forma simpléctica sobre el fibrado cotangente como vemos a continuación.

**Ejemplo 1.1.8.** Sea  $M$  una variedad y sea  $T^*M$  su fibrado cotangente el cual es una variedad  $2n$ -dimensional. considere  $\pi : T^*M \rightarrow M$  la proyección natural, es decir,  $\pi(x, \zeta) = x$  donde  $x \in M$  y  $\zeta \in T_x^*M$ . Definimos puntualmente

$$\alpha_z = (d\pi_z)^* \zeta, \quad (1.4)$$

donde  $z \in T^*M$  y  $z = (x, \zeta)$ . Esta 1-forma se conoce como **forma de Liouville**.

Veamos cómo se ve en coordenadas locales. Sabemos que existe una carta coordenada  $(U, q_1, \dots, q_n)$  centrada en  $x \in M$  y  $(T^*U, q_1, \dots, q_n, p_1, \dots, p_n)$  las coordenadas cotangentes asociadas centradas en  $(x, \zeta)$ . En estas coordenadas podemos escribir  $\zeta = \sum_{i=1}^n p_i dq_i$  entonces

$$\alpha_z = (d\pi_z)^* \zeta_x = (\pi^* \zeta)_z = \left( \sum_{i=1}^n \pi^*(p_i dq_i) \right)_z = \sum_{i=1}^n (p_i \circ \pi) \pi^* dq_i = \sum_{i=1}^n p_i dq_i,$$

donde en la última igualdad usamos el abuso de notación  $\pi^* dq_i = dq_i$ . Con eso vemos que si definimos  $\omega = d\alpha$ , entonces  $\omega_z$  es cerrada y además localmente se ve como

$$\omega = \sum_{i=1}^n dp_i \wedge dq_i$$

que ya vimos que es no degenerada, luego  $\omega$  define una 2-forma simpléctica sobre  $T^*M$ .

Las variedades simplécticas tienen algunos subconjuntos de interés que son isotrópicos, coisotrópicos, simplécticos y lagrangianos. Para nuestro interés solo consideraremos los lagrangianos por su importancia en la mecánica clásica que estudiaremos a continuación.

**Definición 1.1.9.** Sea  $(M, \omega)$  una variedad simpléctica  $2n$ -dimensional. Una subvariedad  $Y$  de  $M$  se llama **subvariedad lagrangiana**, si para cada  $x \in Y$  se tiene  $\omega|_{T_x Y} = 0$  y  $\dim(T_x Y) = \frac{1}{2} \dim(T_x M)$ .

Sea  $i_Y : Y \hookrightarrow M$  la función inclusión. Note que  $Y$  es lagrangiano si y solo si  $i_Y^* \omega = 0$  y  $\dim(Y) = \frac{1}{2} \dim(M)$ . La siguiente proposición nos da una forma de conseguir subvariedades lagrangianas.

**Proposición 1.1.10.** Sean  $(M_1, \omega_1)$  y  $(M_2, \omega_2)$  variedades simplécticas. Un difeomorfismo  $\phi : M_1 \rightarrow M_2$  es **simplectomorfismo** si y solo si  $\Gamma_\phi := \text{Gra}(\phi) = \{(x, \phi(x)) : x \in M_1\}$  es una subvariedad lagrangiana de  $(M_1 \times M_2, \tilde{\Omega})$  donde  $\tilde{\Omega}$  está definido como en el ejemplo [1.1.5](#).

**Demostración:** Sea  $\gamma : M_1 \rightarrow M_1 \times M_2$  definida por  $\gamma(x) = (x, \phi(x))$  con  $x \in M_1$ . Note que

$$\gamma^* \tilde{\Omega} = \gamma^* \pi_1^* \omega_1 - \gamma^* \pi_2^* \omega_2 = (\pi_1 \circ \gamma)^* \omega_1 - (\pi_2 \circ \gamma)^* \omega_2 = \text{Id}_{M_1}^* \omega_1 - \phi^* \omega_2 = \omega_1 - \phi^* \omega_2,$$

por tanto  $\Gamma_\phi$  es lagrangiano  $\iff \gamma^* \tilde{\Omega} = 0 \iff \omega_1 - \phi^* \omega_2 = 0 \iff \phi^* \omega_2 = \omega_1$ ,  $\iff \phi$  es simplectomorfismo. ■

**Definición 1.1.11.** Una **sección** de  $T^*M$  es una función continua  $s : M \rightarrow T^*M$  tal que  $\pi \circ s = \text{Id}_M$  donde  $\pi : T^*M \rightarrow M$  es la función proyección.

## 1.1. UNA INTRODUCCIÓN A LA GEOMETRÍA SIMPLÉCTICA Y DE CONTACTO 7

**Ejemplo 1.1.12.** Considere la sección cero de  $T^*M$  dada por

$$X_0 = \{(x, \zeta) \in T^*M : \zeta = 0 \text{ en } T_x^*M\},$$

la cual es una subvariedad  $n$ -dimensional de  $T^*M$ . Considere una carta coordenada  $(U, q_1, \dots, q_n)$  con su respectiva carta coordenada cotangente  $(T^*U, q_1, \dots, q_n, p_1, \dots, p_n)$ , entonces en  $X_0 \cap T^*U$  se tiene que  $\zeta_1 = \dots = \zeta_n = 0$ , por tanto  $\alpha = 0$  en  $X_0 \cap T^*U$ , luego  $i^*\omega = 0$  y así  $X_0$  es lagrangiano.

Ahora considere las siguientes secciones de  $T^*M$ :

$$X_\mu = \{(x, \mu_x) : x \in M, \mu_x \in T_x^*M\},$$

donde  $\mu : M \rightarrow T^*M$  es una 1-forma y  $\mu_x$  depende suavemente sobre  $x$ .

**Proposición 1.1.13.** Sea  $s_\mu : M \rightarrow T^*M$  definida por  $s_\mu(x) = (x, \mu_x)$ . Si  $\alpha$  es la 1-forma de Liouville de  $T^*M$ , entonces  $s_\mu^*\alpha = \mu$ .

**Demostración:** Sea  $z = s_\mu(x) = (x, \mu_x)$ , luego la 1-forma de Liouville en  $z$  tiene la forma  $\alpha_z = (d\pi_z)^*\mu_x$ , luego,

$$(s_\mu^*\alpha)_x = (ds_\mu)_x^*\alpha_z = (ds_\mu)_x^*((d\pi_z)^*\mu_x) = (d(\pi \circ s_\mu))_x^*\mu_x = (d(\text{id}_M))_x^*\mu_x = \mu_x,$$

por tanto  $s_\mu^*\alpha = \mu$ . ■

Ahora veremos cuando  $X_\mu$  es una subvariedad lagrangiana.

**Proposición 1.1.14.**  $X_\mu$  es una subvariedad lagrangiana de  $T^*M$  si y solo si  $\mu$  es cerrado.

**Demostración:** Considere la función inclusión  $i : X_\mu \rightarrow T^*M$  y la función  $\tau : M \rightarrow X_\mu$  definida como  $\tau(x) = (x, \mu_x)$ , la cual es un difeomorfismo de tal forma que  $s_\mu = i \circ \tau$ . Note que  $X_\mu$  es lagrangiano  $\iff i^*\omega = 0 \iff i^*d\alpha = 0 \iff \tau^*(i^*d\alpha) = 0 \iff (i \circ \tau)^*d\alpha = 0 \iff s_\mu^*d\alpha = 0 \iff d(s_\mu^*\alpha) = 0 \iff d\mu = 0 \iff \mu$  es cerrado. ■

Por tanto, se tiene una correspondencia uno a uno entre el conjunto de subvariedades lagrangianas de  $T^*M$  de la forma  $X_\mu$  y el conjunto de las 1-formas cerradas sobre  $M$ . En el caso que  $M$  es simplemente conexo, por el teorema de Poincaré se tiene que una forma cerrada también es exacta, luego en este caso se puede escribir  $\mu = dF$  para alguna función  $F : M \rightarrow \mathbb{R}$  y en este caso  $F$  se conoce como **función generadora** para la subvariedad lagrangiana  $X_\mu$  asociada a  $\mu$ .

Ahora considere  $M_1$  y  $M_2$  variedades  $n$ -dimensionales con sus respectivos fibrados cotangentes  $T^*M_1$  y  $T^*M_2$  y respectivas 1-formas de Liouville y 2-formas canónicas dadas por:  $\alpha_1, \alpha_2, \omega_1$  y  $\omega_2$ . Entonces como vimos en el ejemplo [1.1.4](#), sobre  $T^*M_1 \times T^*M_2$  se puede definir la 2-forma  $\Omega = \pi_1^*\omega_1 + \pi_2^*\omega_2$  y también la forma torcida dada por  $\tilde{\Omega} = \pi_1^*\omega_1 - \pi_2^*\omega_2$ .

Defina una involución sobre  $T^*M_2$ ,  $\sigma_2 : T^*M_2 \rightarrow T^*M_2$  dada por  $\sigma_2(x_2, \zeta_2) = (x_2, -\zeta_2)$ . Note que  $\sigma_2^*\alpha_2 = \sigma_2^*((d\pi_2)^*\zeta_2) = -(d\pi_2)^*\zeta_2 = -\alpha_2$ . Con esto se puede definir la función  $\sigma := \text{Id}_{M_1} \times \sigma_2 : T^*M_1 \times T^*M_2 \rightarrow T^*M_1 \times T^*M_2$  de la cual se tiene que  $\sigma^*\tilde{\Omega} = \Omega$ .

Ahora tome  $Y$  subvariedad lagrangiana de  $(T^*M_1 \times T^*M_2, \Omega)$  entonces  $Y^\sigma := \sigma(Y)$  es una subvariedad lagrangiana de  $(T^*M_1 \times T^*M_2, \tilde{\Omega})$  y por tanto por la proposición 1.1.10 si  $Y^\sigma$  es la gráfica de algún difeomorfismo  $\phi : T^*M_1 \rightarrow T^*M_2$  entonces  $\phi$  es simplectomorfismo.

Considere  $F \in C^\infty(M_1 \times M_2)$ , luego  $dF$  es una 1-forma en  $M_1 \times M_2$  (en el caso que  $M_1 = M_2$ ,  $F$  se llama una función de dos puntos). Por la proposición 1.1.14 se tiene que  $Y_F = \left\{ \left( (x, y), (dF)_{(x,y)} \right) : (x, y) \in M_1 \times M_2 \right\}$  es un subvariedad lagrangiana de  $T^*(M_1 \times M_2)$  y se llama **subvariedad lagrangiana generada por F**.

Podemos escribir  $Y_F = \{(x, y, d_x F, d_y F) : (x, y) \in M_1 \times M_2\}$  donde  $d_x F := (dF)_{(x,y)}$  proyectado en  $T_x^*M_1 \times \{0\}$  y  $d_y F := (dF)_{(x,y)}$  proyectado en  $\{0\} \times T_x^*M_2$  y por tanto se tiene  $Y_F^\sigma = \{(x, y, d_x F, -d_y F) : (x, y) \in M_1 \times M_2\}$ . En el caso que  $Y_F^\sigma$  es la gráfica de un difeomorfismo  $\phi : T^*M_1 \rightarrow T^*M_2$  entonces  $\phi$  se llama **simplectomorfismo generado por F** y  $F$  se llama **función generadora** de  $\phi$ .

Veamos cuales son las condiciones sobre  $F$  para que genere un simplectomorfismo  $\phi$ . Considere  $(U_1, q_1, \dots, q_n)$  y  $(U_2, p_1, \dots, p_n)$  cartas coordenadas de  $M_1$  y  $M_2$  centradas en  $x$  e  $y$  respectivamente y además  $(T^*U_1, q_1, \dots, q_n, \zeta_1, \dots, \zeta_n)$  y  $(T^*U_2, p_1, \dots, p_n, \eta_1, \dots, \eta_n)$  cartas coordenadas de sus fibrados cotangentes. Entonces  $Y_F^\sigma$  es la gráfica de  $\phi : T^*M_1 \rightarrow T^*M_2 \iff$  para todo  $(x, \zeta) \in T^*M_1$  y  $(y, \eta) \in T^*M_2$  se tiene  $\phi(x, \zeta) = (y, \eta)$  lo cual es equivalente a  $\zeta = d_x F$  y  $\eta = -d_y F$ , por tanto se debe resolver el sistema

$$\begin{cases} \zeta_i = \frac{\partial F}{\partial q_i}(x, y) \\ \eta_i = -\frac{\partial F}{\partial p_i}(x, y). \end{cases} \quad (1.5)$$

Si la primera ecuación tiene solución entonces por el teorema de la función implícita podemos escribir la solución como  $y = \phi_1(x, \zeta)$  siempre y cuando localmente  $F$  cumpla que  $\det \left( \frac{\partial}{\partial p_j} \left( \frac{\partial F}{\partial q_i} \right) \right)_{i,j=1}^n \neq 0$  y al colocar esto en la segunda ecuación se tiene  $\eta = \phi_2(x, \zeta)$  y por tanto  $\phi(x, \zeta) = (\phi_1(x, \zeta), \phi_2(x, \zeta))$ .

### 1.1.2. Geometría de Contacto

El término geometría de contacto apareció por primera vez en 1872 por Sophus Lie quien introdujo la noción de transformación de contacto como una herramienta geométrica para estudiar sistemas de ecuaciones diferenciales [G01]. Sin embargo, las ideas claves se tienen desde Huygens con su formulación de la óptica geométrica [Mc15]. No obstante, esta geometría no recibía tanta atención como su contraparte de dimensión par la geometría simpléctica, pero últimamente se ha trabajado más en esta área dando aplicaciones a la termodinámica, mecánica, teoría de control, entre otros. En esta subsección utilizaremos principalmente las ideas de [dS01] y [Mc15].

**Definición 1.1.15.** *Un elemento de contacto sobre una variedad  $M$   $n$ -dimensional, es un punto  $x \in M$  llamado **punto de contacto**, junto con un hiperplano tangente en  $x$ ,  $H_x \subset T_x M$ , es decir,  $H_x$  es subespacio de  $T_x M$  con  $\dim(H_x) = n - 1$ .*

Note que un hiperplano  $H_x \subset T_x M$  puede ser determinado por el kernel de una 1-forma  $\alpha_x \in T_x^* M - \{0\}$  salvo multiplicación por un escalar, es decir,  $(x, H_x)$  es un elemento

## 1.1. UNA INTRODUCCIÓN A LA GEOMETRÍA SIMPLÉCTICA Y DE CONTACTO 9

de contacto si y solo si  $H_x = \ker(\alpha_x)$  donde  $\alpha_x : T_x M \rightarrow \mathbb{R}$  es lineal y distinta de cero. Además, se puede ver que  $\ker(\alpha_x) = \ker(\alpha'_x) \iff \alpha_x = \lambda \alpha'_x$  para algún  $\lambda \in \mathbb{R} - \{0\}$ .

**Definición 1.1.16.** Una **estructura de contacto** sobre  $M$  es un campo suave de hiperplanos tangentes  $H \subset TM$ , con una 1-forma  $\alpha$  que cumple que  $d\alpha|_H$  es no degenerada (es decir,  $(d\alpha|_H)^n \neq 0$ ) y  $\ker(\alpha) = H$ . La pareja  $(M, H)$  o  $(M, \alpha)$  es llamada una **variedad de contacto** y  $\alpha$  es llamada una **forma local de contacto**.

Note que dado  $x \in M$  podemos escribir  $T_x M = \ker(\alpha_x) \oplus \ker(d\alpha_x)$ . Como  $d\alpha_x|_{H_x}$  es no degenerado tenemos que esta es una forma simpléctica sobre  $H_x$  y por tanto  $\dim(H_x) = 2n$  y como  $\dim(\ker(d\alpha_x)) = 1$  se tiene que  $\dim(M) = 2n + 1$ . Luego las variedades de contacto tienen dimensión impar.

**Proposición 1.1.17.** Sea  $H$  un campo de hiperplanos tangentes sobre  $M$ .  $H$  es una estructura de contacto si y solo si  $\alpha \wedge (d\alpha)^n \neq 0$  para toda 1-forma  $\alpha$  de contacto, es decir,  $H_x = \ker(\alpha_x)$  para  $x \in M$ .

**Demostración:**  $\Rightarrow$  Se tiene por la construcción.

$\Leftarrow$  Suponga que  $H = \ker(\alpha)$ . Considere una base de  $TM = \ker(\alpha) \oplus A$  dada por  $\{e_1, f_1, \dots, e_n, f_n, r\}$  tal que  $\ker(\alpha_x) = \text{Gen}\{e_1, f_1, \dots, e_n, f_n\}$  y  $A = \text{Gen}\{r\}$  luego  $(\alpha \wedge (d\alpha)^n)(e_1, f_1, \dots, e_n, f_n, r) = \alpha(r)(d\alpha)^n(e_1, f_1, \dots, e_n, f_n)$  pero  $\alpha(r) \neq 0$  ya que  $r \notin \ker(\alpha)$  luego  $(d\alpha)^n|_H \neq 0$  y por tanto  $d\alpha|_H$  es no degenerado. ■

Debido a la descomposición de  $T_x M$  existe  $\xi$  sobre  $M$  tal que

$$\iota_\xi d\alpha = 0, \quad (\xi \in \ker(d\alpha)) \quad (1.6)$$

$$\iota_\xi \alpha = 1, \quad (\text{normalización}) \quad (1.7)$$

obtenemos un único vector  $\xi$  que llamamos **campo de Reeb**.

**Proposición 1.1.18.** Sea  $(M, H)$  una variedad de contacto, con forma de contacto  $\alpha$ . El flujo  $\rho_t$  del campo de Reeb  $\xi$  preserva la forma de contacto, es decir,  $\rho_t^* \alpha = \alpha$ .

**Demostración:** Basta con hacer la cuenta

$$\frac{d(\rho_t^* \alpha)}{dt} = \rho_t^* (\mathcal{L}_\xi \alpha) = \rho_t^* (d(\iota_\xi \alpha) + \iota_\xi (d\alpha)) = \rho_t^* (d(1) + 0) = 0,$$

luego  $\rho_t^* \alpha$  es constante y como  $\rho_0 = \text{Id}_M$  se tiene  $\rho_t^* \alpha = \alpha$ . ■

**Ejemplo 1.1.19.** Sea  $M = \mathbb{R}^{2n+1}$  con coordenadas  $(U, q_1, \dots, q_n, p_1, \dots, p_n, S)$ . Entonces la forma

$$\alpha_0 = dS - \sum_{i=1}^n p_i dq_i \quad (1.8)$$

es de contacto. Además, para este caso tenemos que el campo de Reeb es  $\xi = \frac{\partial}{\partial S}$  lo cual se puede comprobar al ver que se cumplen las ecuaciones  $\iota_\xi d\alpha = 0$  y  $\iota_\xi \alpha = 1$ .

**Definición 1.1.20.** Sean  $(M_1, \alpha_1)$  y  $(M_2, \alpha_2)$  variedades de contacto  $(2n+1)$ -dimensionales, y sea  $f : M_1 \rightarrow M_2$  un difeomorfismo. Si  $f^* \alpha_2 = g \alpha_1$  para alguna función  $g : M_1 \rightarrow \mathbb{R}$  entonces  $f$  se llama **contactomorfismo**.

El siguiente teorema es el análogo del teorema de Darboux para el caso simpléctico, así para el caso de una variedad de contacto siempre se puede ver localmente como un espacio que ya es conocido  $\mathbb{R}^{2n+1}$ .

**Teorema 1.1.21.** (Darboux) *Sea  $(M, H)$  una variedad de contacto  $(2n+1)$ -dimensional y  $\alpha$  su forma de contacto. Entonces para todo  $x \in M$  existe una carta coordenada  $(U, q_1, \dots, q_n, p_1, \dots, p_n, S)$ ,  $x \in U$  y un difeomorfismo  $\phi : U \rightarrow \mathbb{R}^{2n+1}$  tal que  $\phi^*\alpha_0 = \alpha$  en  $U$ , donde  $\alpha_0$  está definido por (1.8).*

Con lo anterior parece que hay alguna relación entre variedades simplécticas y variedades de contacto. Vemos que las primeras son de dimensión par y las segundas de dimensión impar. Resulta que podemos pasar de la una a la otra mediante la simplectificación y la contactificación.

**Teorema 1.1.22.** (Simplectificación) *Sea  $(M, H)$  una variedad de contacto de dimensión  $2n-1$  y  $\alpha$  su forma de contacto. Considere  $\widetilde{M} = M \times \mathbb{R}$  y sea  $\pi : \widetilde{M} \rightarrow M$  la proyección  $\pi(x, t) = x$  con  $x \in M$  y  $t \in \mathbb{R}$ . Entonces  $\omega = d(e^t\pi^*\alpha)$  es una forma simpléctica de  $\widetilde{M}$  de dimensión  $2n$ . Entonces  $(\widetilde{M}, \omega)$  se llama la **simplectificación** de  $(M, H)$ .*

De la misma forma tenemos la contactificación.

**Teorema 1.1.23.** (Contactificación) *Sea  $(M, \omega)$  una variedad simpléctica  $2n$ -dimensional. Considere  $\widetilde{M} = M \times \mathbb{R}$  y sea  $\pi : \widetilde{M} \rightarrow M$  la proyección  $\pi(x, t) = x$  con  $x \in M$  y  $t \in \mathbb{R}$ . Si  $\omega$  es exacta, es decir,  $\omega = d\alpha$  con  $\alpha$  una 1-forma sobre  $M$  entonces  $\eta = \pi^*\alpha + dS$  es una forma de contacto de  $\widetilde{M}$  de dimensión  $2n+1$ , donde  $S$  es una coordenada de  $\mathbb{R}$ . Entonces  $(\widetilde{M}, \eta)$  se llama la **contactificación** de  $(M, \omega)$ .*

**Ejemplo 1.1.24.** *Considere  $T^*M$  la cual sabemos que es una variedad simpléctica exacta ( $\omega = d\alpha$ ). Entonces  $T^*M \times \mathbb{R}$  es la contactificación de  $T^*M$  con forma de contacto dada por  $\eta = \pi^*\alpha + dS$  donde  $\alpha$  es la forma de Liouville y  $\pi : T^*M \times \mathbb{R} \rightarrow T^*M$  la proyección.*

Note que  $\omega$  siempre se puede hacer exacta localmente por el teorema de Darboux y lema de Poincaré.

## 1.2. Ecuaciones de Hamilton para Sistemas Conservativos

La segunda ley de Newton nos dice que si tenemos una partícula de masa  $m$  entonces se cumple la relación  $\vec{F} = m\vec{a}$  donde  $\vec{F}$  es la suma de las fuerzas externas sobre la partícula y  $\vec{a}$  su aceleración. Consideremos la coordenada  $q$  como la posición de la partícula en cierto tiempo y además suponga que la fuerza  $\vec{F}$  es conservativa, es decir,  $\vec{F} = -\nabla_q U$  donde  $U$  es una función asociada a la energía potencial de la partícula, entonces la ecuación de Newton queda escrita como:

$$\ddot{q} + \frac{1}{m}\nabla_q U = 0, \quad (1.9)$$

donde la notación  $\ddot{q}$  denota la segunda derivada respecto al tiempo. Veamos cómo podemos comparar esto con el formalismo que estamos usando. Para ello seguiremos las ideas de [Mc15].

**Definición 1.2.1.** Sea  $(M, \omega)$  variedad simpléctica. Sea  $X$  un campo vectorial sobre  $M$ . Cuando  $\mathcal{L}_X \omega = 0$ ,  $X$  es llamado **campo vectorial simpléctico**.

Note que por la identidad de Cartan como  $\omega$  es cerrada se tiene

$$\mathcal{L}_X \omega = d(\iota_X \omega) + \iota_X d\omega = d(\iota_X \omega).$$

Si  $X$  es simpléctico entonces  $d(\iota_X \omega) = 0$ , luego  $\iota_X \omega$  es cerrado.

**Definición 1.2.2.** Si  $\iota_X \omega$  es exacto (lo cual se tiene localmente por el lema de Poincaré) entonces existe  $H : M \rightarrow \mathbb{R}$  tal que

$$\iota_X \omega = -dH.$$

En este caso denotamos  $X = X_H$  y lo llamamos **campo vectorial Hamiltoniano** con función Hamiltoniana  $H$ .

Este campo es único dada  $H$ , ya que existe un isomorfismo entre los campos vectoriales y las 1-formas dado por  $\phi : T_x M \rightarrow T_x^* M$  definido como  $\phi(X) = \iota_X \omega$ .

Como  $\iota_X \omega = -dH$  se tiene  $d(\iota_X \omega) = 0$  y por tanto  $\mathcal{L}_X \omega = 0$  así  $\mathcal{L}_X \Omega = 0$  donde  $\Omega = \frac{(-1)^{\frac{n(n-1)}{2}}}{n!} \omega^n$  y es la forma de volumen de la variedad simpléctica, esto es equivalente a decir que se preserva la forma de volumen, lo cual se conoce como **Teorema de Liouville**.

Para lo que sigue necesitamos la siguiente definición.

**Definición 1.2.3.** Sea  $(M, \omega)$  variedad simpléctica y  $f, g \in C^\infty(M : \mathbb{R})$ . Entonces se define el **Corchete de Poisson**  $\{f, g\} : M \rightarrow \mathbb{R}$  como

$$\{f, g\} := \omega(X_f, X_g). \quad (1.10)$$

Sean  $f, g, h \in C^\infty(M : \mathbb{R})$  entonces algunas propiedades del corchete de Poisson son:

- $\{af + bg, h\} = a\{f, h\} + b\{g, h\}$  con  $a$  y  $b$  constantes (Linealidad).
- $\{h, af + bg\} = a\{h, f\} + b\{h, g\}$  con  $a$  y  $b$  constantes (Linealidad).
- $\{f, g\} = -\{g, f\}$  (Antisimetría).
- $\{fg, h\} = f\{g, h\} + g\{f, h\}$  (Regla de Leibniz).
- $\{f, gh\} = g\{f, h\} + h\{f, g\}$  (Regla de Leibniz).
- $\{\{f, g\}, h\} + \{\{g, h\}, f\} + \{\{h, f\}, g\} = 0$  (Identidad de Jacobi).

Ahora veremos algunas propiedades de los sistemas Hamiltonianos.

**Teorema 1.2.4.** Sea  $(M, \omega)$  variedad simpléctica y  $H$  su función Hamiltoniana.  $\{f, H\} = 0$  si y solo si  $f$  es constante a lo largo de las curvas integrales de  $X_H$ . Esta  $f$  se llama **integral de movimiento** o **primera integral**. En particular,  $H$  es una primera integral, pues  $\{H, H\} = 0$ .

**Demostración:** Sea  $\rho_t$  el flujo de  $X_H$ . Se tiene  $\frac{d}{dt}(f \circ \rho_t) = \rho_t^* \mathcal{L}_{X_H} f = \rho_t^* \iota_{X_H} df = \rho_t^* \iota_{X_H} \iota_{X_f} \omega = \rho_t^* \omega(X_f, X_H) = \rho_t^* \{f, H\}$ , luego  $\{f, H\} = 0 \iff \frac{d}{dt}(f \circ \rho_t) = 0$ . ■



**Proposición 1.2.5.** Si  $f, g \in C^\infty(M : \mathbb{R})$  son primeras integrales de  $X_H$  entonces  $\{f, g\}$  es una primera integral de  $X_H$ .

**Demostración:** Basta con usar la identidad de Jacobi y la antisimetría del corchete de Poisson. Como  $f$  y  $g$  son integrales primeras se tiene que  $\{H, f\} = 0 = \{H, g\}$  luego  $X_H(\{f, g\}) = d\{f, g\}(X_H) = -\iota_{X_{\{f, g\}}}\omega(X_H) = -\omega(X_H, X_{\{f, g\}}) = -\{H, \{f, g\}\} = -\{\{g, H\}, f\} - \{\{H, f\}, g\} = 0$ . ■

A continuación, utilizamos la notación  $[, ]$  para el corchete de Lie, es decir,  $[X, Y](f) = X(Y(f)) - Y(X(f))$  para una función  $f$ .

**Proposición 1.2.6.** Sea  $(M, \omega)$  una variedad simpléctica y  $f, g \in C^\infty(M : \mathbb{R})$  entonces  $X_{\{f, g\}} = [X_f, X_g]$ . Como consecuencia se tiene que si  $X$  y  $Y$  son campos vectoriales Hamiltonianos sobre  $(M, \omega)$ , entonces  $[X, Y]$  es un campo vectorial Hamiltoniano con función Hamiltoniana  $H = \omega(X, Y)$ .

Para la prueba de este resultado ver [Mc15, p. 366 y p. 368].

**Proposición 1.2.7.** Sean  $(M_1, \omega_1)$  y  $(M_2, \omega_2)$  variedades simplécticas y sea  $\phi : M_1 \rightarrow M_2$  un simplectomorfismo y  $f \in C^\infty(M_2 : \mathbb{R})$  entonces  $\phi^*X_f = X_{f \circ \phi}$ .

**Demostración:** Sea  $Y$  campo vectorial sobre  $M_2$  entonces  $(\iota_{\phi^*X_f}\omega_1)(Y) = (\iota_{\phi^*X_f}\phi^*\omega_2)(Y) = \phi^*\omega_2(\phi^*X_f, Y) = \omega_2(\phi_*\phi^*X_f, \phi_*Y) = \omega_2(X_f, \phi_*Y) = \iota_{X_f}\omega_2(\phi_*Y) = -df(\phi_*Y) = -\phi^*(df(Y)) = -d(\phi^*f(Y)) = -d(f \circ \phi)(Y)$  luego se tiene que  $\iota_{\phi^*X_f}\omega_2 = -d(f \circ \phi)$  y por tanto  $\phi^*X_f = X_{f \circ \phi}$ . ■

**Proposición 1.2.8.** Sean  $(M_1, \omega_1)$  y  $(M_2, \omega_2)$  variedades simplécticas y sea  $\phi : M_1 \rightarrow M_2$  un difeomorfismo.  $\phi$  es simplectomorfismo si y solo si para toda  $f, g \in C^\infty(M_2 : \mathbb{R})$  se tiene  $\{f \circ \phi, g \circ \phi\} = \{f, g\} \circ \phi$ .

Para la demostración ver [Mc15, p. 370].

**Definición 1.2.9.** Un **álgebra de Lie** es un espacio vectorial  $\mathcal{G}$  junto con un corchete de Lie  $[, ] : \mathcal{G} \times \mathcal{G} \rightarrow \mathcal{G}$  que es bilineal, antisimétrico y que satisface la identidad de Jacobi. Se denota como  $(\mathcal{G}, [, ])$ .

Así si denotamos  $\chi^{Ham}$  como los campos vectoriales Hamiltonianos sobre  $M$ ,  $\chi^{Sim}$  los campos vectoriales simplécticos sobre  $M$  y  $\chi$  los campos vectoriales sobre  $M$  se tiene la siguiente contención entre álgebras de Lie:

$$(\chi^{Ham}, [, ]) \subset (\chi^{Sim}, [, ]) \subset (\chi, [, ])$$

Ahora veremos qué forma tiene  $X_H$  en coordenadas de Darboux. Para  $x \in M$  escoga un sistema de coordenadas  $(U, q_1, \dots, q_n, p_1, \dots, p_n)$  y una función  $H : M \rightarrow \mathbb{R}$ ,  $H(q_1, \dots, q_n, p_1, \dots, p_n)$  luego se tiene  $dH = \sum_{j=1}^n \left( \frac{\partial H}{\partial q_j} dq_j + \frac{\partial H}{\partial p_j} dp_j \right)$ . Además considere

$$X_H = \sum_{j=1}^n \left( X_{q_j} \frac{\partial}{\partial q_j} + X_{p_j} \frac{\partial}{\partial p_j} \right) \text{ y } Y = \sum_{j=1}^n \left( Y_{q_j} \frac{\partial}{\partial q_j} + Y_{p_j} \frac{\partial}{\partial p_j} \right), \text{ luego se tiene:}$$



$$\begin{aligned}
\iota_{X_H}\omega_0(Y) &= \omega_0(X_H, Y) \\
&= \sum_{j=1}^n (dp_j \wedge dq_j)(X_H, Y) \\
&= \sum_{j=1}^n (dp_j(X_H) dq_j(Y) - dq_j(X_H) dp_j(Y)) \\
&= \sum_{j=1}^n (X_{p_j} dq_j(Y) - X_{q_j} dp_j(Y)) \\
&= \sum_{j=1}^n (X_{p_j} dq_j - X_{q_j} dp_j)(Y),
\end{aligned}$$

por tanto  $\iota_{X_H}\omega_0 = \sum_{j=1}^n (X_{p_j} dq_j - X_{q_j} dp_j)$ , así la condición  $\iota_{X_H}\omega_0 = -dH$  se convierte en

$$\sum_{j=1}^n (X_{p_j} dq_j - X_{q_j} dp_j) = \sum_{j=1}^n \left( -\frac{\partial H}{\partial q_j} dq_j - \frac{\partial H}{\partial p_j} dp_j \right),$$

de donde se obtiene que  $X_{p_j} = -\frac{\partial H}{\partial q_j}$  y  $X_{q_j} = \frac{\partial H}{\partial p_j}$  para  $j \in \{1, \dots, n\}$ . Luego  $X_H$  tiene la forma

$$X_H = \sum_{j=1}^n \left( \frac{\partial H}{\partial p_j} \frac{\partial}{\partial q_j} - \frac{\partial H}{\partial q_j} \frac{\partial}{\partial p_j} \right). \quad (1.11)$$

Si consideramos las curvas integrales de  $X_H$ ,  $\lambda(t) = (q_1(t), \dots, q_n(t), p_1(t), \dots, p_n(t))$ , entonces se obtiene para todo  $j \in \{1, \dots, n\}$  las ecuaciones de Hamilton,

$$\begin{cases} \dot{q}_j = \frac{\partial H}{\partial p_j} \\ \dot{p}_j = -\frac{\partial H}{\partial q_j}. \end{cases} \quad (1.12)$$

Además note que  $\dot{H} = \sum_{j=1}^n \left( \frac{\partial H}{\partial q_j} \dot{q}_j + \frac{\partial H}{\partial p_j} \dot{p}_j \right) = 0$  luego  $H$  es constante a lo largo de las curvas integrales de  $X_H$ . Lo que confirma lo dicho en teorema [1.2.4](#)

**Ejemplo 1.2.10.** Considere una partícula con posición  $q = (q_1, \dots, q_n)$  y cantidad de movimiento  $p = (p_1, \dots, p_n)$ . Tomemos un Hamiltoniano de la forma  $H = K + U$  donde  $K$  es la energía cinética de la partícula dada por  $K = \frac{\|p\|^2}{2m}$  y  $U$  su energía potencial que depende de  $q$ . Entonces la segunda ecuación de Hamilton toma la forma  $\dot{p}_i = -\frac{\partial H}{\partial q_i} = -\frac{\partial U}{\partial q_i}$ . Si la masa es constante se tiene  $p_i = m\dot{q}_i$  luego  $m\ddot{q}_i + \frac{\partial U}{\partial q_i} = 0$  la cual tiene la forma de [\(1.9\)](#), de aquí que las ecuaciones [\(1.12\)](#) se llamen ecuaciones de Hamilton para sistemas conservativos donde  $H$  es la función de energía la cual se conserva en el tiempo pues  $\dot{H} = 0$ .

Si se considera la carta coordenada  $(U, q_1, \dots, q_n, p_1, \dots, p_n)$  y la ecuación [\(1.11\)](#) para  $f$  y  $g$  podemos escribir el corchete de Poisson en coordenadas de Darboux como:

$$\{f, g\}_{(q_k, p_k)} = \sum_{k=1}^n \left( \frac{\partial f}{\partial q_k} \frac{\partial g}{\partial p_k} - \frac{\partial f}{\partial p_k} \frac{\partial g}{\partial q_k} \right),$$

luego las ecuaciones de Hamilton se pueden escribir como:

$$\begin{cases} \dot{q}_i = \{q_i, H\}_{(q_k, p_k)}, \\ \dot{p}_i = \{p_i, H\}_{(q_k, p_k)}. \end{cases}$$

Ahora definiremos las transformaciones canónicas y nos guiaremos de [BCT17].

**Definición 1.2.11.** Sea  $(M, \omega)$  una variedad simpléctica. Una **transformación canónica** es un automorfismo  $\phi : M \rightarrow M$  que preserva la 2-forma simpléctica  $\omega$ .

En el ejemplo 1.1.8 vimos que  $T^*M$  es una variedad  $2n$ -dimensional con una 2-forma  $\omega$  asociada. Sea  $U \subset M$  un abierto con coordenadas de Darboux  $(T^*U, q_1, \dots, q_n, p_1, \dots, p_n)$  y  $H$  el Hamiltoniano, luego  $\omega = \sum_{i=1}^n dp_i \wedge dq_i$ , además se satisfacen las ecuaciones (1.12).

Ahora sea  $V \subset M$  un abierto con coordenadas de Darboux  $(T^*V, Q_1, \dots, Q_n, P_1, \dots, P_n)$  y tome  $K$  el Hamiltoniano en estas coordenadas. Si existe una transformación canónica entonces se preserva la 2-forma simpléctica y así  $\phi^*\omega = \sum_{i=1}^n dP_i \wedge dQ_i$  es la 2-forma en estas coordenadas y por tanto  $K$  satisface las ecuaciones (1.12) escritas en estas coordenadas como  $\dot{Q}_i = \frac{\partial K}{\partial P_i}$  y  $\dot{P}_i = -\frac{\partial K}{\partial Q_i}$  y por tanto se tiene que  $K(Q, P) = H(q(Q, P), p(Q, P))$ . De aquí vemos que las transformaciones canónicas dejan invariantes las ecuaciones de Hamilton.

Como se preserva la 2-forma entonces la transformación canónica debe dejar invariante la 1-forma canónica salvo por un diferencial exacto, luego se tiene que

$$\sum_{i=1}^n p_i dq_i = \sum_{i=1}^n P_i dQ_i + dF. \quad (1.13)$$

La función  $F$  es llamada **función generadora**, la cual puede ser de 4 tipos distintos dependiendo de las coordenadas de las que depende, que pueden ser  $F_1(q_i, Q_i)$  (tipo I),  $F_2(q_i, P_i)$  (tipo II),  $F_3(p_i, Q_i)$  (tipo III) o  $F_4(p_i, P_i)$  (tipo IV).

Ahora tomaremos  $F_1(q_1, \dots, q_n, Q_1, \dots, Q_n)$  la cual es llamada **función generadora de tipo I** de la transformación canónica. Note que  $dF_1 = \sum_{i=1}^n \left( \frac{\partial F_1}{\partial q_i} dq_i + \frac{\partial F_1}{\partial Q_i} dQ_i \right)$  si reemplazamos en la nueva forma se tiene

$$\sum_{i=1}^n p_i dq_i = \sum_{i=1}^n P_i dQ_i + \sum_{i=1}^n \left( \frac{\partial F_1}{\partial q_i} dq_i + \frac{\partial F_1}{\partial Q_i} dQ_i \right) = \sum_{i=1}^n \left( \frac{\partial F_1}{\partial q_i} dq_i + \left( P_i + \frac{\partial F_1}{\partial Q_i} \right) dQ_i \right),$$

y de aquí se obtiene que para  $i \in \{1, \dots, n\}$ ,

$$\begin{cases} p_i = \frac{\partial F_1}{\partial q_i} \\ P_i = -\frac{\partial F_1}{\partial Q_i}. \end{cases} \quad (1.14)$$

Estas son las condiciones que cumple  $F_1$  si la transformación es canónica. Note que estas condiciones son las mismas que se obtuvieron en la ecuación (1.5) para la función generadora de las subvariedades lagrangianas.

La siguiente proposición nos da una forma de verificar si una transformación es canónica.

**Proposición 1.2.12.** *Considere la transformación que manda  $(U, q_1, \dots, q_n, p_1, \dots, p_n)$  a  $(V, Q_1, \dots, Q_n, P_1, \dots, P_n)$  tales que*

$$\begin{cases} \{Q_i, Q_j\}_{(q_k, p_k)} = 0, \\ \{P_i, P_j\}_{(q_k, p_k)} = 0, \\ \{Q_i, P_j\}_{(q_k, p_k)} = \delta_{i,j}, \end{cases}$$

entonces la transformación es canónica.

**Ejemplo 1.2.13.** *El Hamiltoniano para el oscilador armónico 1-dimensional está dado por  $H = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2$ . Sean las nuevas coordenadas*

$$\begin{aligned} q(Q, P) &= \sqrt{\frac{2P}{m\omega}} \operatorname{sen} Q, \\ p(Q, P) &= \sqrt{2m\omega P} \cos Q. \end{aligned}$$

Note que

$$\{q, p\}_{(Q, P)} = \sqrt{\frac{2P}{m\omega}} \cos Q \sqrt{\frac{m\omega}{2P}} \cos Q - \frac{1}{\sqrt{2m\omega P}} \operatorname{sen} Q \left( -\sqrt{2m\omega P} \operatorname{sen} Q \right) = 1.$$

Se puede ver que  $\{q, q\} = 0$  y  $\{p, p\} = 0$  luego esta es una transformación canónica. Así el Hamiltoniano en las nuevas coordenadas toma la forma  $K(Q, P) = \omega P$ . Además, la función generadora de tipo I es  $F_1(q, Q) = \frac{m\omega q^2}{2} \cot Q$ . En efecto, veamos que se cumplen las ecuaciones (1.14).

$$\frac{\partial F_1}{\partial q} = m\omega q \cot Q = m\omega q \frac{p}{m\omega q} = p.$$

$$\begin{aligned} \frac{\partial F_1}{\partial Q} &= -\frac{m\omega q^2}{2} \operatorname{csc}^2 Q \\ &= -\frac{m\omega q^2}{2} (1 + \cot^2 Q) \\ &= -\frac{m\omega q^2}{2} \left( 1 + \frac{p^2}{m^2 \omega^2 q^2} \right) \\ &= -\frac{H}{\omega} = -\frac{K}{\omega} = -P. \end{aligned}$$

La función generadora puede ser obtenida solucionando las ecuaciones (1.14).

### 1.3. Ecuaciones de Hamilton para Sistemas Disipativos

Sabemos que en la naturaleza no es común tener sistemas que conserven la energía, es decir, que cumplan  $\dot{H} = 0$ . Además, en optimización, vamos a querer utilizar la función energía  $H$  como una función de Lyapunov para probar la convergencia. Por eso necesitamos encontrar métodos matemáticos para considerar estos casos. En las siguientes subsecciones consideraremos 3 formas de realizar disipación de la energía geoméricamente y nos guiaremos de [BCT17].

### 1.3.1. Hamiltonianos dependientes del tiempo

Nuestro primer caso de disipación de energía será guiado de la referencia [BCT17] y vamos a considerar un Hamiltoniano dependiente del tiempo. Para ello consideraremos el espacio de fase extendido  $T^*M \times \mathbb{R}$ . Sea  $\pi : T^*M \times \mathbb{R} \rightarrow T^*M$  la función proyección y  $\alpha$  la 1-forma de Liouville del Ejemplo 1.1.8. Entonces  $T^*M \times \mathbb{R}$  es una variedad  $(2n + 1)$ -dimensional y además es una variedad de contacto donde le asignamos la 1-forma dada por

$$\alpha_{PC} = \pi^*\alpha - H dt, \quad (1.15)$$

la cual se llama **1-forma de Poincare-Cartan**, donde  $H$  es el Hamiltoniano que puede o no depender del tiempo. Note que esto no es la contactificación del Teorema 1.1.23

**Proposición 1.3.1.**  $\alpha_{PC}$  es una forma de contacto.

*Demostración.* Note que  $d\alpha_{PC} = d(\pi^*\alpha) - dH \wedge dt$  y por tanto se tiene

$$(d\alpha_{PC})^n = (d\pi^*\alpha)^n - n(d\pi^*\alpha)^{n-1} \wedge dH \wedge dt, \quad (1.16)$$

luego,

$$\alpha_{PC} \wedge (d\alpha_{PC})^n = \pi^*\alpha \wedge (d\pi^*\alpha)^n - n\pi^*\alpha \wedge (d\pi^*\alpha)^{n-1} \wedge dH \wedge dt - H dt \wedge (d\pi^*\alpha)^n.$$

Primero note que  $\pi^*\alpha \wedge (d\pi^*\alpha)^n = 0$  pues es una  $(2n + 1)$ -forma que tiene las coordenadas (de Darboux)  $q_1, \dots, q_n, p_1, \dots, p_n$ . Por otro lado, vamos a calcular el término  $\pi^*\alpha \wedge (d\pi^*\alpha)^{n-1} \wedge dH \wedge dt$ . Para ello usaremos coordenadas de Darboux  $(q, p)$ , en las cuales se tiene  $dH = \sum_{j=1}^n \left( \frac{\partial H}{\partial q_j} dq_j + \frac{\partial H}{\partial p_j} dp_j \right) + \frac{\partial H}{\partial t} dt$  y  $\pi^*\alpha = \sum_{j=1}^n p_j dq_j$  y por tanto

$$(d\pi^*\alpha)^{n-1} = (n-1)! \sum_{j=1}^n dp_1 \wedge dq_1 \wedge \dots \wedge dp_{j-1} \wedge dq_{j-1} \wedge \hat{d}p_j \wedge \hat{d}q_j \wedge dp_{j+1} \wedge dq_{j+1} \wedge \dots \wedge dp_n \wedge dq_n,$$

$$(d\pi^*\alpha)^n = n! dp_1 \wedge dq_1 \wedge \dots \wedge dp_n \wedge dq_n,$$

donde la notación  $\hat{d}p_j \wedge \hat{d}q_j$  significa que esos términos no están incluidos en el producto. Al hacer la multiplicación correspondiente se tiene que

$$\begin{aligned} n\pi^*\alpha \wedge (d\pi^*\alpha)^{n-1} \wedge dH \wedge dt &= n! \sum_{j=1}^n p_j \frac{\partial H}{\partial p_j} dp_1 \wedge dq_1 \wedge \dots \wedge dp_n \wedge dq_n \wedge dt \\ &= - \sum_{j=1}^n p_j \frac{\partial H}{\partial p_j} dt \wedge (d\pi^*\alpha)^n, \end{aligned}$$

Luego se obtiene,

$$\alpha_{PC} \wedge (d\alpha_{PC})^n = \left( \sum_{j=1}^n p_j \frac{\partial H}{\partial p_j} - H \right) dt \wedge (d\pi^*\alpha)^n.$$

Sabemos que  $(d\pi^*\alpha)^n \neq 0$  y además si  $\sum_{j=1}^n p_j \frac{\partial H}{\partial p_j} \neq H$  se tiene que  $\alpha_{PC} \wedge (d\alpha_{PC})^n \neq 0$ . Así  $\alpha_{PC}$  es una forma de contacto. ■

Como  $\alpha_{PC}$  es una forma de contacto, podemos escribirla en coordenadas  $(q, p)$  como:

$$\alpha_{PC} = \sum_{i=1}^n p_i dq_i - H dt.$$

Con esto podemos calcular las ecuaciones del movimiento y para ello consideramos  $X_H = \sum_{j=1}^n \left( X_{q_j} \frac{\partial}{\partial q_j} + X_{p_j} \frac{\partial}{\partial p_j} \right) + X_t \frac{\partial}{\partial t}$ . Un cálculo directo nos permite ver que  $\iota_{X_H^E} d\alpha_{PC} = 0$  si y solo si  $X_H^E = X_t \left( X_H + \frac{\partial}{\partial t} \right)$  donde  $X_H$  esta dado por la ecuación (1.11). Por simplicidad consideraremos  $X_t = 1$  de donde se obtiene que  $\dot{t} = 1$  y las ecuaciones de movimiento son:

$$\begin{cases} \dot{q}_j &= \frac{\partial H}{\partial p_j} \\ \dot{p}_j &= -\frac{\partial H}{\partial q_j} \\ \dot{t} &= 1. \end{cases} \quad (1.17)$$

Note que  $\dot{H} = \sum_{j=1}^n \left( \frac{\partial H}{\partial q_j} \dot{q}_j + \frac{\partial H}{\partial p_j} \dot{p}_j \right) + \frac{\partial H}{\partial t} \dot{t} = \frac{\partial H}{\partial t}$  y por tanto vemos que la energía no se conserva si  $\frac{\partial H}{\partial t} \neq 0$ . Además, las ecuaciones son las mismas que las de Hamilton (1.12).

Ahora veremos que se tiene una transformación canónica para el sistema Hamiltoniano dependiente del tiempo con su respectiva función generadora.

Considere la carta coordenada  $(U, q_1, \dots, q_n, p_1, \dots, p_n, t)$  y sea el cambio de coordenadas  $(V, Q_1, \dots, Q_n, P_1, \dots, P_n, t)$ , así siguiendo la idea para las transformaciones canónicas para el caso de sistemas Hamiltonianos conservativos, se tiene la condición:

$$\sum_{i=1}^n p_i dq_i - H dt = \left( \sum_{i=1}^n P_i dQ_i - K dt \right) + dF, \quad (1.18)$$

donde  $K : T^*M^E \rightarrow \mathbb{R}$  es el nuevo Hamiltoniano y  $F : T^*M^E \rightarrow \mathbb{R}$  es la función generadora. Tomando el caso en que las coordenadas para  $F$  son  $(q_1, \dots, q_n, Q_1, \dots, Q_n, t)$  (tipo I) entonces la ecuación (1.18) se puede escribir como:

$$\sum_{i=1}^n \left( \left( p_i - \frac{\partial F}{\partial q_i} \right) dq_i - \left( P_i + \frac{\partial F}{\partial Q_i} \right) dQ_i \right) + \left( K - H - \frac{\partial F}{\partial t} \right) dt = 0,$$

de aquí la función generadora para esta transformación satisface:

$$\begin{cases} p_i - \frac{\partial F}{\partial q_i} = 0 \\ P_i + \frac{\partial F}{\partial Q_i} = 0 \\ K - H - \frac{\partial F}{\partial t} = 0, \end{cases} \quad (1.19)$$

así las ecuaciones de Hamilton en las nuevas coordenadas pueden ser escritas como en (1.12) con una ecuación adicional como sigue,

$$\begin{cases} \dot{Q}_i &= \frac{\partial K}{\partial P_i} \\ \dot{P}_i &= -\frac{\partial K}{\partial Q_i} \\ \dot{t} &= 1. \end{cases}$$

Un caso especial de transformación canónica es la que lleva a  $K = 0$ . Es este caso, escribiendo  $F = S$  y usando la tercera ecuación de (1.19) se obtiene la **ecuación de Hamilton-Jacobi**

$$H\left(q_1, \dots, q_n, \frac{\partial S}{\partial q_1}, \dots, \frac{\partial S}{\partial q_n}, t\right) = -\frac{\partial S}{\partial t}.$$

La solución  $S(q_1, \dots, q_n, t)$  se llama **función principal de Hamilton** la cual determina completamente la dinámica del sistema dado en (1.12) con la ecuación adicional  $\dot{t} = 1$  y con ciertas condiciones iniciales. Note que hemos reducido las ecuaciones a una sola ecuación diferencial parcial cuya solución tiene toda la información necesaria para obtener las trayectorias del sistema mecánico.

### 1.3.2. Hamiltonianos conformalmente simplécticos

Como siguiente caso de disipación de energía consideraremos los sistemas Hamiltonianos conformes y para ello nos guiaremos de las referencias [McP01] y [FSRV19].

**Definición 1.3.2.** Sea  $(M, \omega)$  una variedad simpléctica. Un campo vectorial  $X^c$  se dice **conforme** con parámetro  $c \in \mathbb{R}$  si  $\mathcal{L}_{X^c}\omega = c\omega$ . Además  $\phi^C$  se llama **difeomorfismo conforme** si  $(\phi^C)^*\omega = C\omega$  con  $C$  una constante.

Considere la siguiente relación

$$\iota_{X^c}\omega = -dH - c\alpha, \quad (1.20)$$

donde  $\alpha$  es la 1-forma de Liouville y por tanto  $\omega = -d\alpha$ . Note que esta relación satisface la condición de ser conforme. En efecto, utilizando la identidad de Cartan se tiene

$$\mathcal{L}_{X^c}\omega = d(\iota_{X^c}\omega) + \iota_{X^c}(d\omega) = d(-dH - c\alpha) = c\omega,$$

lo que satisface la Definición 1.3.2. Realizando un cálculo usando (1.20) se puede ver que un campo vectorial conforme satisface las ecuaciones:

$$\begin{cases} \dot{q}_i = \frac{\partial H}{\partial p_i} \\ \dot{p}_i = -\frac{\partial H}{\partial q_i} - cp_i, \end{cases} \quad (1.21)$$

donde  $c > 0$  es una constante de amortiguamiento responsable de la disipación de la energía del sistema. Además, utilizando (1.21) se tiene que  $\dot{H} = -c \sum_{i=1}^n \frac{\partial H}{\partial p_i} p_i$  lo cual muestra que no se conserva la energía.

Por otro lado, note que el sistema (1.21) puede ser escrito como

$$\dot{z} = \Omega \nabla H(z) - cDz, \quad (1.22)$$

donde

$$z = \begin{pmatrix} q \\ p \end{pmatrix}, \quad \Omega = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 0 \\ 0 & I \end{pmatrix},$$

La ecuación (1.22) define un flujo  $\Phi_t : M \rightarrow M$  tal que  $\Phi_t(z_0) = z(t)$  con  $z(t)$  la trayectoria en el tiempo  $t$  con condición inicial  $z(0) = z_0$ . Además, note que

$$\frac{d\Phi_t^*\omega}{dt} = \Phi_t^*\mathcal{L}_{X^c}\omega = c\Phi_t^*\omega,$$

solucionando esta ecuación diferencial se tiene  $\omega_t = \Phi_t^*\omega = e^{ct}\omega$ .

### 1.3.3. Hamiltonianos de contacto

Para nuestro tercer ejemplo de disipación de energía seguiremos las ideas de [BCT17].

**Definición 1.3.3.** Sea  $(M, H)$  una variedad de contacto  $(2n + 1)$ -dimensional con  $\alpha$  su 1-forma que define localmente a  $H$ .  $X$  se llama **campo vectorial de contacto** si  $X$  es un campo vectorial tal que su flujo  $\phi_t$  es un contactomorfismo para todo  $t$ . En otras palabras,  $\mathcal{L}_X \alpha = g\alpha$  para alguna función  $g : M \rightarrow \mathbb{R}$ . Si  $g = 0$  el campo vectorial se llama **campo vectorial estrictamente de contacto**.

**Definición 1.3.4.** Sea  $(M, H)$  una variedad de contacto  $(2n + 1)$ -dimensional con  $\alpha$  su 1-forma que define localmente a  $H$ . Sea  $\mathcal{H} : M \rightarrow \mathbb{R}$ , entonces  $X_{\mathcal{H}}$  se llama **campo vectorial Hamiltoniano generado por  $\mathcal{H}$**  si se cumplen las ecuaciones,

$$\mathcal{L}_{X_{\mathcal{H}}} \alpha = f_{\mathcal{H}} \alpha \quad (1.23)$$

$$-\mathcal{H} = \iota_{X_{\mathcal{H}}} \alpha, \quad (1.24)$$

donde  $f_{\mathcal{H}} : M \rightarrow \mathbb{R}$  y  $\mathcal{H}$  se llama **Hamiltoniano de contacto**.

Note que aunque la estructura de contacto no depende de  $\alpha$ , el campo Hamiltoniano sí depende de esta 1-forma de contacto, luego a partir de ahora consideraremos un  $\alpha$  fijo y definido globalmente. Ahora vamos a encontrar expresiones para  $\mathcal{H}$  y  $X_{\mathcal{H}}$  en coordenadas de Darboux.

**Lema 1.3.5.** Sea  $(M, H)$  una variedad de contacto  $(2n + 1)$ -dimensional con  $\alpha$  su 1-forma de contacto. Entonces la función  $f_{\mathcal{H}}$  de la definición 1.3.4 tiene la forma  $f_{\mathcal{H}} = -\xi(\mathcal{H})$ . En particular en coordenadas de Darboux  $f_{\mathcal{H}} = -\frac{\partial \mathcal{H}}{\partial S}$ .

*Demostración.* Considere la identidad de Cartan

$$\mathcal{L}_{X_{\mathcal{H}}} \alpha = \iota_{X_{\mathcal{H}}} (d\alpha) + d(\iota_{X_{\mathcal{H}}} \alpha) = \iota_{X_{\mathcal{H}}} (d\alpha) - d\mathcal{H}.$$

Sea  $\xi$  el campo de Reeb asociado a  $\alpha$  y note que

$$\begin{aligned} \iota_{\xi} (\mathcal{L}_{X_{\mathcal{H}}} \alpha) &= \iota_{\xi} (\iota_{X_{\mathcal{H}}} (d\alpha) - d\mathcal{H}) \\ &= \iota_{\xi} (\iota_{X_{\mathcal{H}}} (d\alpha)) - \iota_{\xi} (d\mathcal{H}) \\ &= -\iota_{\xi} (d\mathcal{H}) = -d\mathcal{H}(\xi) = -\xi(\mathcal{H}), \end{aligned}$$

donde  $\iota_{\xi} (\iota_{X_{\mathcal{H}}} (d\alpha)) = 0$  por ser el campo de Reeb. Por otro lado  $\iota_{\xi} (\mathcal{L}_{X_{\mathcal{H}}} \alpha) = \iota_{\xi} (f_{\mathcal{H}} \alpha) = f_{\mathcal{H}}$  pues  $\iota_{\xi} \alpha = 1$  por ser campo de Reeb. Por tanto, juntando esto se obtiene que  $f_{\mathcal{H}} = -\xi(\mathcal{H}) = -\frac{\partial \mathcal{H}}{\partial S}$ , la última igualdad se tiene porque en una carta de Darboux se tiene que  $\xi = \frac{\partial}{\partial S}$ . ■

Utilizando las fórmulas de la definición de campo Hamiltoniano de contacto se puede obtener mediante un cálculo parecido al realizado para el caso de  $X_H$  en el caso simpléctico que

$$X_{\mathcal{H}} = \sum_{i=1}^n \left( \frac{\partial \mathcal{H}}{\partial p_i} \frac{\partial}{\partial q_i} - \left( p_i \frac{\partial \mathcal{H}}{\partial S} + \frac{\partial \mathcal{H}}{\partial q_i} \right) \frac{\partial}{\partial p_i} + \left( p_i \frac{\partial \mathcal{H}}{\partial p_i} - \mathcal{H} \right) \frac{\partial}{\partial S} \right), \quad (1.25)$$

de esta forma las ecuaciones del movimiento vienen dadas por:

$$\begin{cases} \dot{q}_i = \frac{\partial \mathcal{H}}{\partial p_i} \\ \dot{p}_i = -\frac{\partial \mathcal{H}}{\partial q_i} - p_i \frac{\partial \mathcal{H}}{\partial S} \\ \dot{S} = \sum_{i=1}^n p_i \frac{\partial \mathcal{H}}{\partial p_i} - \mathcal{H}, \end{cases} \quad (1.26)$$

las cuales podemos ver que son una generalización de las ecuaciones de Hamilton (1.12) para una variedad de contacto. Note que si  $\mathcal{H}$  no depende de  $S$  las primeras dos ecuaciones se reducen a las ecuaciones de Hamilton (1.12). Sabemos que la energía total de un sistema es  $H = \frac{\|p\|^2}{2} + V(q)$  con  $V$  la función potencial. Tomando derivadas respecto al tiempo, usando (1.26) y que  $\mathcal{H} = H + g(S)$ , donde  $g$  es una función de  $S$ , se tiene

$$\begin{aligned} \dot{H} &= \sum_{j=1}^n \left( \frac{\partial H}{\partial q_j} \dot{q}_j + \frac{\partial H}{\partial p_j} \dot{p}_j \right) \\ &= \sum_{j=1}^n \left( \frac{\partial H}{\partial q_j} \frac{\partial \mathcal{H}}{\partial p_j} + \frac{\partial H}{\partial p_j} \left( -\frac{\partial \mathcal{H}}{\partial q_j} - p_j \frac{\partial \mathcal{H}}{\partial S} \right) \right) \\ &= -\sum_{j=1}^n p_j \frac{\partial H}{\partial p_j} \frac{\partial \mathcal{H}}{\partial S} \\ &= -\|p\|^2 \frac{\partial \mathcal{H}}{\partial S} \\ &= -\|p\|^2 \frac{\partial g(S)}{\partial S}, \end{aligned}$$

luego la energía total del sistema no se conserva. Además, si  $\frac{\partial g(S)}{\partial S} > 0$  se tiene que la energía del sistema siempre decrece, lo cual nos da una guía para construir sistemas Hamiltonianos de contacto que tengan a la función de energía total como función de Lyapunov.

Al igual que en el caso simpléctico, tenemos transformaciones de contacto las cuales preservan la forma de contacto salvo por multiplicación por una función  $f : M \rightarrow \mathbb{R} - \{0\}$ , es decir,  $\tilde{\alpha} = f\alpha$ , así si consideramos la carta coordenada  $(U, q_1, \dots, q_n, p_1, \dots, p_n, S)$  y queremos cambiarla a la carta  $(V, Q_1, \dots, Q_n, P_1, \dots, P_n, \tilde{S})$  entonces una transformación de contacto está dada por la condición:

$$f \left( dS - \sum_{i=1}^n p_i dq_i \right) = d\tilde{S} - \sum_{i=1}^n P_i dQ_i.$$

Escribiendo el lado derecho se tiene,

$$f dS - \sum_{i=1}^n f p_i dq_i = \sum_{i,j=1}^n \left( \frac{\partial \tilde{S}}{\partial q_j} - P_i \frac{\partial Q_i}{\partial q_j} \right) dq_j + \sum_{i,j=1}^n \left( \frac{\partial \tilde{S}}{\partial p_j} - P_i \frac{\partial Q_i}{\partial p_j} \right) dp_j + \sum_{i=1}^n \left( \frac{\partial \tilde{S}}{\partial S} - P_i \frac{\partial Q_i}{\partial S} \right) dS,$$

de aquí obtenemos las condiciones:



$$\begin{cases} -fp_i = \sum_{j=1}^n \left( \frac{\partial \tilde{S}}{\partial q_i} - P_j \frac{\partial Q_j}{\partial q_i} \right) \\ 0 = \sum_{j=1}^n \left( \frac{\partial \tilde{S}}{\partial p_i} - P_j \frac{\partial Q_j}{\partial p_i} \right) \\ f = \sum_{j=1}^n \left( \frac{\partial \tilde{S}}{\partial S} - P_j \frac{\partial Q_j}{\partial S} \right). \end{cases} \quad (1.27)$$

Si el Hamiltoniano depende del tiempo debemos hacer la extensión  $M^E = M \times \mathbb{R}$  que es una variedad  $(2n + 2)$ -dimensional, donde  $M = T^*N \times \mathbb{R}$  con  $N$  una variedad  $n$ -dimensional.  $M^E$  es llamada **variedad de contacto extendida**. Ahora extendemos la 1-forma  $\alpha$  como,

$$\alpha^E = dS - \pi^* \alpha + \mathcal{H}dt,$$

donde  $\mathcal{H}$  es el Hamiltoniano de contacto y dependerá del tiempo y  $\alpha$  es la 1-forma de Liouville.

**Proposición 1.3.6.**  $(M^E, d\alpha^E)$  es un espacio simpléctico<sup>1</sup>.

*Demostración.* Note que  $d\alpha^E = d\mathcal{H} \wedge dt - d(\pi^* \alpha)$  luego se tiene

$$(d\alpha^E)^{n+1} = (n+1)(d\pi^* \alpha)^n \wedge d\mathcal{H} \wedge dt - (d\pi^* \alpha)^{n+1} = (n+1)(d\pi^* \alpha)^n \wedge d\mathcal{H} \wedge dt,$$

pues  $(d\pi^* \alpha)^{n+1} = 0$  ya que es una  $(2n + 2)$ -forma que solo depende de las coordenadas  $q_1, \dots, q_n, p_1, \dots, p_n$ . Además, en coordenadas de Darboux se tiene que

$$(d\pi^* \alpha)^n \wedge d\mathcal{H} \wedge dt = n! \frac{\partial \mathcal{H}}{\partial S} dp_1 \wedge \dots \wedge dp_n \wedge dq_1 \wedge \dots \wedge dq_n \wedge dS \wedge dt,$$

luego  $(d\alpha^E)^{n+1} \neq 0$  siempre y cuando  $\frac{\partial \mathcal{H}}{\partial S} \neq 0$ . Por tanto  $d\alpha^E$  define una 2-forma simpléctica. ■

Podemos ver que en coordenadas toma la forma:

$$\alpha^E = dS - \sum_{i=1}^n p_i dq_i + \mathcal{H}dt.$$

Para definir la dinámica en  $M^E$  tenemos las condiciones

$$\mathcal{L}_{X_{\mathcal{H}}^E} \alpha^E = g_{\mathcal{H}} \alpha^E \quad (1.28)$$

$$\iota_{X_{\mathcal{H}}^E} \alpha^E = 0, \quad (1.29)$$

donde  $g_{\mathcal{H}} : M^E \rightarrow \mathbb{R}$ . Si usamos la fórmula de Cartan con las condiciones anteriores se tiene:

$$g_{\mathcal{H}} \alpha^E = \mathcal{L}_{X_{\mathcal{H}}^E} \alpha^E = d \left( \iota_{X_{\mathcal{H}}^E} \alpha^E \right) + \iota_{X_{\mathcal{H}}^E} (d\alpha^E) = \iota_{X_{\mathcal{H}}^E} (d\alpha^E).$$

Usando estas condiciones podemos encontrar mediante el proceso hecho anteriormente para  $X_H$  que:

$$X_{\mathcal{H}}^E = X_{\mathcal{H}} + \frac{\partial}{\partial t},$$

con  $X_{\mathcal{H}}$  dado por la ecuación (1.25), por tanto, las ecuaciones de movimiento son las mismas que en el caso de contacto (1.26) con la ecuación adicional  $\dot{t} = 1$ .

<sup>1</sup>Note que esto no es la simplectificación del Teorema 1.1.22.

**Ejemplo 1.3.7.** Considere el Hamiltoniano  $\mathcal{H} = \sum_{i=1}^n \frac{p_i^2}{2m} + V(q, t) + f(t)S$ . Usando las ecuaciones (1.26) se tiene

$$\begin{cases} \dot{q}_i = \frac{p_i}{m} \\ \dot{p}_i = -\frac{\partial V}{\partial q_i} - f(t)p_i \\ \dot{S} = \sum_{i=1}^n \frac{p_i^2}{2m} - V(q, t) - f(t)S. \end{cases}$$

Usando las dos primeras ecuaciones y la notación  $q = (q_1, \dots, q_n)$  se obtiene la ecuación

$$m\ddot{q} + f(t)\dot{q} + \nabla_q V = 0,$$

la cual corresponde a la segunda ley de Newton con una fuerza de fricción lineal en la velocidad y dependiente del tiempo, es decir, un sistema no conservativo. Si  $f(t) = 0$  recuperamos el caso conservativo.

También se tiene transformaciones de contacto dependientes del tiempo. Luego considere la carta coordenada  $(U, q_1, \dots, q_n, p_1, \dots, p_n, S, t)$  y queremos cambiarla a la carta  $(V, Q_1, \dots, Q_n, P_1, \dots, P_n, \tilde{S}, t)$  entonces debemos tener la condición:

$$f \left( dS - \sum_{i=1}^n p_i dq_i + \mathcal{H} dt \right) = d\tilde{S} - \sum_{i=1}^n P_i dQ_i + \mathcal{K} dt,$$

donde  $\mathcal{K}$  es el Hamiltoniano de contacto en las nuevas coordenadas. Colocando los respectivos diferenciales de  $dQ_i$  y  $d\tilde{S}$  se obtienen las mismas condiciones que (1.27) con la ecuación adicional  $f\mathcal{H} = \frac{\partial \tilde{S}}{\partial t} - \sum_{i=1}^n P_i \frac{\partial Q_i}{\partial t} + \mathcal{K}$ .

La geometría de la información utiliza la geometría diferencial para estudiar problemas en estadística, ciencia de datos, Machine Learning y otros campos relacionados. Harold Hotelling fue el primero en considerar la matriz de información de Fisher como un tensor métrico Riemanniano, de aquí el trabajo de muchos matemáticos ha continuado explorando esta rama, por mencionar algunos se tiene Harold Jeffreys, Solomon Kullback, Calyampudi Radhakrishna Rao, Nikolai Chentsov, Shun'ichi Amari, entre muchos otros (Ver [N18] y [A16]).

## 2.1. Variedades Estadísticas

Vamos a considerar  $(M, g)$  una variedad con una métrica  $g$ . Esta sección utiliza las ideas de [N18] y [L19].

**Definición 2.1.1.** Una conexión  $\nabla^*$  se dice **conjugada** a una conexión  $\nabla$  con respecto a un tensor métrico  $g$  si y solo si para todo  $X, Y, Z \in \chi(M)$  se satisface:

$$Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z). \quad (2.1)$$

Lo primero que podemos notar es que  $(\nabla^*)^* = \nabla$ . En efecto, como  $(\nabla^*)^*$  es la conexión conjugada de  $\nabla^*$  se tiene que:

$$\begin{aligned} Xg(Y, Z) &= g(\nabla_X^* Y, Z) + g(Y, (\nabla^*)^*_X Z) \\ &= Xg(Y, Z) - g(Y, \nabla_X Z) + g(Y, (\nabla^*)^*_X Z), \end{aligned}$$

donde en la segunda igualdad se despejo de la ecuación (2.1) y de aquí se tiene que  $g(Y, \nabla_X Z) = g(Y, (\nabla^*)^*_X Z)$ , luego  $(\nabla^*)^* = \nabla$ . Note que en el caso Riemanniano se tiene  $\nabla = \nabla^*$  la cual combinada con que  $\nabla$  sea libre de torsión, es decir,  $\nabla_X Y - \nabla_Y X = [X, Y]$ , es la conexión de Levi-Civita. A partir de aquí asumiremos que  $\nabla$  y  $\nabla^*$  son libres de torsión.

**Definición 2.1.2.**  $(M, g, \nabla, \nabla^*)$  denota una **variedad con conexiones conjugadas** donde  $(\nabla, \nabla^*)$  son conexiones conjugadas respecto a  $g$ .

**Proposición 2.1.3.** Dada  $\nabla$  en  $(M, g)$  existe una única conexión conjugada  $\nabla^*$ .

*Demostración.* Sean  $\nabla^*$  y  $\tilde{\nabla}^*$  conexiones conjugadas de  $\nabla$  luego se tienen las relaciones  $Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z)$  y  $Xg(Y, Z) = g(\nabla_X Y, Z) + g(Y, \tilde{\nabla}_X^* Z)$  de aquí se tiene que  $g(Y, \nabla_X^* Z) = g(Y, \tilde{\nabla}_X^* Z)$  y por tanto  $\nabla^* = \tilde{\nabla}^*$ . ■

**Proposición 2.1.4.** Sea  $\bar{\nabla} = \frac{\nabla + \nabla^*}{2}$  la cual llamamos **conexión promedio**. Entonces  $\bar{\nabla}$  es autoadjunta, es decir  $\bar{\nabla} = \bar{\nabla}^*$  y además coincide con la conexión métrica de Levi-Civita.

*Demostración.* Tenemos que ver que  $Xg(Y, Z) = g(\bar{\nabla}_X Y, Z) + g(Y, \bar{\nabla}_X Z)$ . En efecto,

$$\begin{aligned} g(\bar{\nabla}_X Y, Z) + g(Y, \bar{\nabla}_X Z) &= g\left(\frac{\nabla_X Y + \nabla_X^* Y}{2}, Z\right) + g\left(Y, \frac{\nabla_X Z + \nabla_X^* Z}{2}\right) \\ &= \frac{1}{2}[g(\nabla_X Y, Z) + g(Y, \nabla_X^* Z)] + \frac{1}{2}[g(\nabla_X^* Y, Z) + g(Y, \nabla_X Z)] \\ &= \frac{1}{2}Xg(Y, Z) + \frac{1}{2}Xg(Z, Y) = Xg(Y, Z). \end{aligned}$$

Además, como  $\nabla$  y  $\nabla^*$  son libres de torsión entonces  $\bar{\nabla}$  es libre de torsión y así  $\bar{\nabla}$  es la conexión de Levi-Civita. ■

**Definición 2.1.5.** Una **variedad estadística**  $(M, g, C)$  es una variedad equipada con un tensor métrico  $g$  y un  $(0, 3)$ -tensor totalmente simétrico llamado el tensor de **Amari-Chentsov** definido como  $C = \nabla g$ , donde  $\nabla$  es una conexión definida sobre  $M$ .

Considere la base  $\{\partial_1, \dots, \partial_n\}$ , entonces el tensor de Amari-Chentsov tiene la forma:

$$\nabla g(\partial_i, \partial_j, \partial_k) = \partial_k g(\partial_i, \partial_j) - g(\nabla_{\partial_k} \partial_i, \partial_j) - g(\partial_i, \nabla_{\partial_k} \partial_j) = \partial_k g_{ij} - \Gamma_{ki,j} - \Gamma_{kj,i}, \quad (2.2)$$

donde estamos usando que en coordenadas locales  $g_{ij} = g(\partial_i, \partial_j)$  y los símbolos de Christoffel  $g(\nabla_{\partial_i} \partial_j, \partial_k) = \Gamma_{ij,k}$  y de manera análoga se definen los símbolos de Christoffel para la conexión conjugada,  $g(\nabla_{\partial_i}^* \partial_j, \partial_k) = \Gamma_{ij,k}^*$ . Ahora usando la ecuación (2.1) se tiene que

$$\partial_k g_{ij} = g(\nabla_{\partial_k} \partial_i, \partial_j) + g(\partial_i, \nabla_{\partial_k}^* \partial_j) = \Gamma_{ki,j} + \Gamma_{kj,i}^*. \quad (2.3)$$

Reemplazando esto en (2.2) se tiene

$$C_{ijk} = \Gamma_{kj,i}^* - \Gamma_{kj,i}, \quad (2.4)$$

o en notación libre de coordenadas para  $X, Y, Z \in \chi(M)$  se puede escribir como:

$$C(X, Y, Z) = g(\nabla_X^* Y - \nabla_X Y, Z). \quad (2.5)$$

Veamos que  $C$  es un tensor totalmente simétrico, es decir,  $C_{ijk} = C_{\sigma(i)\sigma(j)\sigma(k)}$  para toda permutación  $\sigma$ . Como  $\nabla$  y  $\nabla^*$  son libres de torsión se tiene  $\Gamma_{kj,i} = \Gamma_{jk,i}$  y  $\Gamma_{kj,i}^* = \Gamma_{jk,i}^*$ . Luego por (2.4) vemos que  $C_{ijk} = C_{ikj}$  y de la ecuación (2.2) se tiene  $C_{ijk} = C_{jik}$  y así  $C$  es totalmente simétrico. Note que si  $\nabla$  es la conexión de Levi-Civita se tiene que  $\Gamma_{ij,k} = \Gamma_{ij,k}^*$  y por tanto  $C = 0$ . Además, note que haciendo lo mismo con la conexión conjugada  $\nabla^*$  se tiene  $\nabla^* g = -C$ .

**Definición 2.1.6.** Sobre una variedad estadística definimos una **familia uniparamétrica de conexiones afines**  $\nabla^{(\alpha)}$  llamadas  $\alpha$ -conexiones con  $\alpha \in \mathbb{R}$  definidas como:

$$\nabla^{(\alpha)} = \frac{1 + \alpha}{2} \nabla + \frac{1 - \alpha}{2} \nabla^*.$$

Si  $\alpha = 0$  se tiene que  $\nabla^{(0)} = \bar{\nabla}$  la cual es la conexión de Levi-Civita. Si  $\alpha = 1$  se tiene  $\nabla^{(1)} = \nabla$  y para  $\alpha = -1$  se tiene  $\nabla^{(-1)} = \nabla^*$ . Además, realizando un cálculo análogo a la demostración de la proposición [2.1.4](#) se puede ver que  $\nabla^{(\alpha)} = (\nabla^{(-\alpha)})^*$ . Además, el tensor de Amari-Chentsov para la variedad estadística con las conexiones conjugadas  $(\nabla^{(\alpha)}, \nabla^{(-\alpha)})$  está dado por  $C^{(\alpha)} = \alpha C$ . En efecto,

$$C^{(\alpha)} = \nabla^{(\alpha)} g = \frac{1 + \alpha}{2} \nabla g + \frac{1 - \alpha}{2} \nabla^* g = \frac{1 + \alpha}{2} C - \frac{1 - \alpha}{2} C = \alpha C.$$

Ahora para ver los símbolos de Christoffel sea  $X, Y, Z \in \chi(M)$ , entonces

$$g\left(\nabla_X^{(\alpha)} Y, Z\right) = \frac{1 + \alpha}{2} g(\nabla_X Y, Z) + \frac{1 - \alpha}{2} g(\nabla_X^* Y, Z) \quad (2.6)$$

$$= g(\bar{\nabla}_X Y, Z) + \frac{\alpha}{2} C(X, Y, Z). \quad (2.7)$$

De [\(2.7\)](#) se tiene  $\Gamma_{ij,k}^{(\alpha)} = \bar{\Gamma}_{ij,k} + \frac{\alpha}{2} C_{ijk}$ , donde  $\bar{\Gamma}_{ij,k}$  son los símbolos de Christoffel para la conexión de Levi-Civita. Además, por [\(2.6\)](#) vemos que  $\Gamma_{ij,k}^{(\alpha)} = \frac{1 + \alpha}{2} \Gamma_{ij,k} + \frac{1 - \alpha}{2} \Gamma_{ij,k}^*$ , luego tenemos dos formas equivalentes de escribir los símbolos de Christoffel para esta variedad. Así  $(M, g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$  define una variedad de conexiones conjugadas para todo  $\alpha \in \mathbb{R}$ .

## 2.2. Funciones de Divergencia

Una función de divergencia representa un grado de separación entre dos puntos. Sin embargo, no es una distancia pues no cumple la propiedad de simetría, ni la desigualdad triangular. Comenzaremos dando su definición y veremos como esta nos permite inducir una geometría sobre una variedad. Las ideas de esta sección serán tomadas de [\[L19\]](#), [\[N18\]](#), [\[A16\]](#) y [\[LZ17\]](#).

**Definición 2.2.1.** Sea  $M$  una variedad y  $D : M \times M \rightarrow [0, \infty)$ .  $D$  se llama **función de divergencia** si es una función  $C^3$  y además satisface:

- (a)  $D(x, y) \geq 0$  para todo  $x, y \in M$ , con igualdad cuando  $x = y$ .
- (b)  $\partial_{x_i} D(x, y)|_{x=y} = 0 = \partial_{y_j} D(x, y)|_{x=y}$  para todo  $i, j \in \{1, \dots, n\}$ .
- (c)  $-\partial_{x_i} \partial_{y_j} D(x, y)|_{x=y}$  es definida positiva.

Ahora veremos como una función de divergencia es capaz de inducir una métrica Riemanniana  $g$  con un par de conexiones conjugadas libres de torsión.

**Teorema 2.2.2.** *Sea  $M$  una variedad  $n$ -dimensional y sea  $D$  una función de divergencia sobre  $M \times M$ , entonces  $(M, {}^Dg, {}^D\nabla, {}^D\nabla^*)$  es una variedad de conexiones conjugadas donde la métrica y las conexiones están dadas por:*

$${}^Dg_{ij}(x) = -\partial_{x_i}\partial_{y_j}D(x, y)|_{x=y} \quad (2.8)$$

$${}^D\Gamma_{ij,k}(x) = -\partial_{y_k}\partial_{x_i}\partial_{x_j}D(x, y)|_{x=y} \quad (2.9)$$

$${}^D\Gamma_{ij,k}^*(x) = -\partial_{x_k}\partial_{y_i}\partial_{y_j}D(x, y)|_{x=y}. \quad (2.10)$$

Luego  $(M, {}^Dg, {}^DC)$  define una variedad estadística.

*Demostración.* Primero veamos como transforma  ${}^Dg_{ij}$  bajo un cambio de coordenadas. Considere el cambio de coordenadas  $x' = x'(x)$  y  $y' = y'(y)$  entonces

$$\begin{aligned} g_{hl}(x') &= -\partial_{x'_h}\partial_{y'_l}D(x', y')|_{x'=y'} \\ &= -\frac{\partial x'_h}{\partial x_i}\partial_{x_i}\left(\frac{\partial y'_l}{\partial y_j}\partial_{y_j}D(x'(x), y'(y))\right)\Big|_{x'=y'} \\ &= -\frac{\partial x'_h}{\partial x_i}\frac{\partial y'_l}{\partial y_j}\partial_{x_i}\partial_{y_j}D(x'(x), y'(y))|_{x'=y'} - \frac{\partial x'_h}{\partial x_i}\frac{\partial^2 y'_l}{\partial x_i\partial y_j}(\partial_{y_j}D(x'(x), y'(y)))\Big|_{x'=y'} \\ &= \frac{\partial x'_h}{\partial x_i}\frac{\partial y'_l}{\partial y_j}g_{ij}(x), \end{aligned}$$

donde en la tercera igualdad se utilizó el inciso (b) de la definición de divergencia. Además, como  $-\partial_{x_i}\partial_{y_j}D(x, y)|_{x=y}$  es definida positiva por definición de función de divergencia y además es simétrica porque  $D$  es  $C^3$  entonces  ${}^Dg_{ij}$  es una métrica Riemanniana. Para ver que las conexiones están bien definidas es suficiente con ver que se cumple la ecuación (2.3) (dualidad con respecto a  $g$ )

$$\begin{aligned} {}^D\Gamma_{ki,j} + {}^D\Gamma_{kj,i}^* &= -\partial_{y_j}\partial_{x_k}\partial_{x_i}D(x, y)|_{x=y} - \partial_{x_i}\partial_{y_k}\partial_{y_j}D(x, y)|_{x=y} \\ &= (\partial_{x_k} + \partial_{y_k})(-\partial_{x_i}\partial_{y_j}D(x, y))|_{x=y} \\ &= \partial_k {}^Dg_{ij}, \end{aligned}$$

donde  $\partial_k = (\partial_{x_k} + \partial_{y_k})|_{x=y}$ . Además como  $D$  es  $C^3$  entonces  ${}^D\Gamma_{ki,j}$  y  ${}^D\Gamma_{kj,i}^*$  son libres de torsión y el tensor de Amari-Chentsov es  ${}^DC_{ijk} = {}^D\Gamma_{kj,i}^* - {}^D\Gamma_{ki,j}$ . Además, un cálculo parecido al hecho con la métrica muestra que  ${}^D\Gamma_{ki,j}$  y  ${}^D\Gamma_{kj,i}^*$  transforman como símbolos de Christoffel. Así  $(M, {}^Dg, {}^DC)$  es variedad estadística. ■

Como un primer ejemplo veremos la geometría de Bregman.

**Ejemplo 2.2.3.** *Considere  $F : \Omega \rightarrow \mathbb{R}$  una función suave y convexa donde  $\Omega \subset \mathbb{R}^n$  es un conjunto convexo y abierto. Defina la función **divergencia de Bregman** como:*

$$B_F(x, y) = F(x) - F(y) - \langle x - y, \nabla F(y) \rangle, \quad (2.11)$$

donde  $\langle, \rangle$  denota el producto interno usual de  $\mathbb{R}^n$ . Veamos que en efecto es una función de divergencia:

(a) *Como la función es estrictamente convexa entonces se tiene que  $F(x) - F(y) \geq \langle x - y, \nabla F(y) \rangle$  luego  $B_F(x, y) \geq 0$  con igualdad cuando  $x = y$ .*

(b) Tomando la derivada respecto a la primera entrada se tiene:

$$\partial_{x_i} B_F(x, y) = \partial_{x_i} F(x) - \langle \partial_{x_i} x, \nabla F(y) \rangle = \partial_{x_i} F(x) - \partial_{y_i} F(y), \quad (2.12)$$

luego si  $x = y$  se tiene  $\partial_{x_i} B_F(x, x) = 0$ . Por otro lado, tomando la derivada respecto a la segunda entrada se tiene,

$$\begin{aligned} \partial_{y_i} B_F(x, y) &= -\partial_{y_i} F(y) - \langle \partial_{y_i}(-y), \nabla F(y) \rangle - \langle x - y, \partial_{y_i} \nabla F(y) \rangle \\ &= -\partial_{y_i} F(y) + \partial_{y_i} F(y) - \langle x - y, \partial_{y_i} \nabla F(y) \rangle \\ &= -\langle x - y, \partial_{y_i} \nabla F(y) \rangle, \end{aligned} \quad (2.13)$$

de aquí cuando  $x = y$  se tiene  $\partial_{x_i} B_F(x, x) = 0$ .

(c) Con los cálculos anteriores se tiene que

$$\partial_{y_j} \partial_{x_i} B_F(x, y) = \partial_{y_j} (\partial_{x_i} F(x) - \partial_{y_i} F(y)) = -\partial_{y_j} \partial_{y_i} F(y).$$

Como  $F$  es estrictamente convexa entonces  $\partial_{y_j} \partial_{y_i} F(y) \geq 0$  luego cuando  $x = y$  se tiene  $-\partial_{x_j} \partial_{x_i} B_F(x, x) = \partial_{x_j} \partial_{x_i} F(x)$  la cual es una matriz definida positiva.

Así  $B_F$  es una función de divergencia. Ahora podemos ver la estructura de variedad estadística que induce  $B_F$ . El Teorema [2.2.2](#) y los cálculos realizados arriba nos dice que

$${}^D g_{ij}(x) = \partial_{x_j} \partial_{x_i} F(x),$$

veamos ahora los símbolos de Christoffel:

$${}^{B_F} \Gamma_{ij,k} = \partial_{y_k} \partial_{x_i} \partial_{x_j} B_F(x, y)|_{x=y} = \partial_{x_i} (\partial_{y_k} \partial_{y_j} F(y))|_{x=y} = 0,$$

$${}^{B_F} \Gamma_{ij,k}^* = \partial_{x_k} \partial_{y_i} \partial_{y_j} B_F(x, y)|_{x=y} = \partial_{x_k} \partial_{x_i} \partial_{x_j} F(x).$$

La estructura obtenida acá se conoce como **variedad Hessiana**, pues su métrica está dada por la Hessiana de la función  $F$ .

Ahora considere  $F^* : V^* \subset (\mathbb{R}^n)^* \rightarrow \mathbb{R}$  el **dual de Legendre** de  $F$  definido como:

$$F^*(v) := \sup_{x \in V} \{ \langle x, v \rangle - F(x) \}. \quad (2.14)$$

Note que de la definición se tiene que

$$F^*(v) \geq \langle x, v \rangle - F(x), \quad (2.15)$$

la cual se llama **desigualdad de Fenchel** donde  $x \in V$  y  $v \in V^*$ . Veamos cuando se cumple la igualdad. Considere la función  $\tilde{F}(x, v) = \langle x, v \rangle - F(x)$  y vamos a encontrar el máximo de esta función. El máximo se da si  $0 = \nabla_x \tilde{F}(x, v) = v - \nabla F(x^*)$ , donde  $x^*$  es el punto donde se alcanza el máximo. Llame  $w = \nabla F(x^*)$  y así el parámetro dual es  $x^* = (\nabla F)^{-1}(w)$ , así se obtiene

$$F^*(w) = \langle (\nabla F)^{-1}(w), w \rangle - F((\nabla F)^{-1}(w)). \quad (2.16)$$

Esta última expresión es conocida como **transformación de Legendre**. Por otro lado, se tiene que  $(F^*)^*(x) = \sup_{v \in V^*} \{\langle x, v \rangle - F^*(v)\}$  y haciendo lo mismo que antes para encontrar el máximo se tiene que  $x = \nabla F^*(v^*)$  con  $v^*$  el punto donde se alcanza el máximo, luego  $v^* = (\nabla F^*)^{-1}(x)$ . Así la igualdad en (2.15) se tiene cuando  $x$  es la conjugada de Legendre de  $w$  y viceversa, y las llamaremos **variables conjugadas** dadas por lo siguiente:

$$x = \nabla F^*(w) = (\nabla F)^{-1}(w) \longleftrightarrow w = \nabla F(x) = (\nabla F^*)^{-1}(x). \quad (2.17)$$

Además vemos que para todo  $x \in V$  y  $v \in V^*$  se tienen las relaciones  $x = \nabla F^*(\nabla F(x))$  y  $v = \nabla F(\nabla F^*(v))$ , es decir,  $(\nabla F^*)^{-1} = \nabla F$  y  $(\nabla F)^{-1} = \nabla F^*$ .

Veamos que  $(F^*)^* = F$ . En efecto,

$$\begin{aligned} (F^*)^*(x) &= \langle (\nabla F^*)^{-1}(x), x \rangle - F^*((\nabla F^*)^{-1}(x)) \\ &= \langle \nabla F(x), x \rangle - F^*(\nabla F(x)) \\ &= \langle \nabla F(x), x \rangle - \langle (\nabla F)^{-1}(\nabla F(x)), \nabla F(x) \rangle + F((\nabla F)^{-1}(\nabla F(x))) \\ &= \langle \nabla F(x), x \rangle - \langle x, \nabla F(x) \rangle + F(x) \\ &= F(x), \end{aligned}$$

donde en la segunda igualdad se usó el hecho de que  $(\nabla F^*)^{-1} = \nabla F$  y en la primera y tercera igualdad se usó (2.16).

**Definición 2.2.4.** Sean  $u, v \in V^*$ . Se define el **dual de la divergencia de Bregman** como

$$B_{F^*}(u, v) = B_F(\nabla F(u), \nabla F(v)), \quad (2.18)$$

Además, ahora podemos definir otra divergencia en términos de las variables conjugadas (2.17).

**Ejemplo 2.2.5.** Definimos la **divergencia canónica**  $A_F : V \times V^* \rightarrow \mathbb{R}$  como

$$A_F(x, v) = F(x) + F^*(v) - \langle x, v \rangle. \quad (2.19)$$

Note que  $A_F(x, v) = A_{F^*}(v, x)$ .

Veamos que  $A_F$  es una función de divergencia.

(a) Por la desigualdad de Fenchel se tiene que  $A_F(x, v) \geq 0$  con igualdad dada por (2.17).

(b) Derivando (2.19) respecto a la primera y segunda entrada se tiene

$$\partial_{x_i} A_F(x, v) = \partial_{x_i} F(x) - v_i, \quad (2.20)$$

$$\partial_{v_i} A_F(x, v) = \partial_{v_i} F^*(v) - x_i, \quad (2.21)$$

nuevamente usando (2.17) se tiene que las derivadas son cero.

(c)  $\partial_{v_j} \partial_{x_i} A_F(x, v) = -\delta_{i,j} = \partial_{x_j} \partial_{v_i} A_F(x, v)$ , luego  $-\partial_{v_i} \partial_{x_i} A_F(x, v)$  es definida positiva.

En este caso tenemos que la métrica está dada por  $g_{ij} = \delta_{i,j}$  y al tomar derivadas nuevamente y usando el Teorema 2.2.2 se tiene que los símbolos de Cristhoffel de las conexiones duales son 0, lo cual me dice que tengo una variedad que es dualmente plana. Además, en términos de la divergencia de Bregman, la divergencia canónica se puede escribir como  $B_F(x, (\nabla F)^{-1}(v)) = A_F(x, v)$ .



## 2.3. Información de Fisher

Considere  $(\Omega, \mathcal{B}, P)$  un espacio de probabilidad, donde  $\Omega$  es el espacio muestral,  $\mathcal{B}$  es una  $\sigma$ -álgebra sobre  $\Omega$  y  $P$  una función de probabilidad. Además, sea  $p : \Omega \rightarrow [0, \infty)$  la función densidad de probabilidad, es decir, dado  $A \subset \Omega$  un conjunto medible se tiene que  $P(A) = \int_A p(x) dx$ . Ahora considere una familia paramétrica de distribuciones de probabilidad

$$\mathcal{P} = \{p_\theta(x) : \theta = (\theta_1, \dots, \theta_n) \in M\},$$

donde  $M$  una subvariedad abierta de  $\mathbb{R}^n$  que llamamos **espacio de parámetros** y  $\theta$  se llama el **parámetro de la familia**. Además, definimos la **función log-verosimilitud** como

$$l(\theta, x) = \ln(p_\theta(x)), \quad (2.22)$$

esta escritura será útil más adelante para los cálculos.

**Definición 2.3.1.** Sean  $p_\theta$  y  $p_\eta$  densidades de probabilidad con  $\theta, \eta \in M$  entonces definimos la función de **entropía relativa** o **divergencia de Kullback-Leibler** entre dos densidades de probabilidad como:

$$D(\theta, \eta) = \int_{\Omega} p_\theta(x) \ln \left( \frac{p_\theta(x)}{p_\eta(x)} \right) dx = E_\theta \left[ \ln \left( \frac{p_\theta(x)}{p_\eta(x)} \right) \right], \quad (2.23)$$

donde  $E_\theta$  es la esperanza respecto a  $p_\theta$ . Además para todo  $x \in \Omega$  tal que  $p_\eta(x) = 0$  se tiene  $p_\theta(x) = 0$ . Cuando  $p_\theta(x) = 0$  la contribución de término correspondiente se toma como cero, pues  $\lim_{x \rightarrow 0^+} x \ln(x) = 0$ .

**Proposición 2.3.2.** La función de entropía relativa (2.23) es una función de divergencia.

*Demostración.* Veamos que (2.23) cumple los incisos de la definición 2.2.1.

(a) La desigualdad de Jensen en formulación probabilista es  $\varphi(E[X]) \leq E[\varphi(X)]$  con  $\varphi$  una función convexa. Tome  $\varphi(x) = -\ln(x)$  la cual es una función convexa luego se tiene:

$$\begin{aligned} E_\theta \left[ \ln \left( \frac{p_\theta(x)}{p_\eta(x)} \right) \right] &= E_\theta \left[ -\ln \left( \frac{p_\eta(x)}{p_\theta(x)} \right) \right] \\ &\geq -\ln \left( E_\theta \left[ \frac{p_\eta(x)}{p_\theta(x)} \right] \right) \\ &= -\ln \left( \int_{\Omega} p_\theta(x) \frac{p_\eta(x)}{p_\theta(x)} dx \right) \\ &= -\ln \left( \int_{\Omega} p_\eta(x) dx \right) = -\ln(1) = 0, \end{aligned}$$

es decir,  $D(\theta, \eta) \geq 0$  además si  $\theta = \eta$  se tiene  $p_\theta = p_\eta$ , luego  $\ln \left( \frac{p_\theta(x)}{p_\eta(x)} \right) = \ln(1) = 0$  y así  $D(\theta, \theta) = 0$ .

(b) Note que la ecuación (2.23) se puede escribir como

$$D(\theta, \eta) = \int_{\Omega} (p_{\theta}(x) \ln(p_{\theta}(x)) - p_{\theta}(x) \ln(p_{\eta}(x))) dx.$$

Calculando la derivada  $\partial_{\theta_i}$  se tiene:

$$\partial_{\theta_i} D(\theta, \eta) = \int_{\Omega} \left( (\partial_{\theta_i} p_{\theta}(x)) \ln(p_{\theta}(x)) + p_{\theta}(x) \frac{1}{p_{\theta}(x)} \partial_{\theta_i} p_{\theta}(x) - (\partial_{\theta_i} p_{\theta}(x)) \ln(p_{\eta}(x)) \right) dx.$$

Si  $\theta = \eta$  el primer y último término se cancelan y se tiene:

$$\partial_{\theta_i} D(\theta, \theta) = \int_{\Omega} \partial_{\theta_i} p_{\theta}(x) dx = \partial_{\theta_i} \int_{\Omega} p_{\theta}(x) dx = \partial_{\theta_i}(1) = 0.$$

Calculando la derivada  $\partial_{\eta_i}$  se tiene:

$$\partial_{\eta_i} D(\theta, \eta) = - \int_{\Omega} p_{\theta}(x) \partial_{\eta_i} \ln(p_{\eta}(x)) dx = - \int_{\Omega} p_{\theta}(x) \frac{1}{p_{\eta}(x)} \partial_{\eta_i} p_{\eta}(x) dx.$$

Si  $\theta = \eta$  se tiene:

$$\partial_{\theta_i} D(\theta, \theta) = - \int_{\Omega} \partial_{\theta_i} p_{\theta}(x) dx = - \partial_{\theta_i} \int_{\Omega} p_{\theta}(x) dx = - \partial_{\theta_i}(1) = 0.$$

(c) Con los cálculos anteriores se puede ver que

$$\begin{aligned} \partial_{\theta_i} \partial_{\eta_j} D(\theta, \eta) &= - \int_{\Omega} \partial_{\theta_i} p_{\theta}(x) \partial_{\eta_j} \ln(p_{\eta}(x)) dx \\ &= - \int_{\Omega} p_{\theta}(x) \frac{\partial_{\theta_i} p_{\theta}(x)}{p_{\theta}(x)} \partial_{\eta_j} \ln(p_{\eta}(x)) dx \\ &= - \int_{\Omega} p_{\theta}(x) \partial_{\theta_i} \ln(p_{\theta}(x)) \partial_{\eta_j} \ln(p_{\eta}(x)) dx \\ &= - \int_{\Omega} p_{\theta}(x) \partial_{\theta_i} l(\theta, x) \partial_{\eta_j} l(\eta, x) dx. \end{aligned}$$

Cuando  $\theta = \eta$  se tiene

$$-\partial_{\theta_i} \partial_{\theta_j} D(\theta, \theta) = \int_{\Omega} p_{\theta}(x) \partial_{\theta_i} l(\theta, x) \partial_{\theta_j} l(\theta, x) dx = E_{\theta} [\partial_{\theta_i} l(\theta, x) \partial_{\theta_j} l(\theta, x)].$$

Esta última expresión sabemos que es definida positiva y por tanto  $D$  es función de divergencia. ■

La última expresión obtenida en la prueba anterior es una matriz bien conocida que veremos a continuación.

**Definición 2.3.3.** La *Matriz de información de Fisher* se define como

$$\mathcal{I}_{ij}(\theta) = E_{\theta} [\partial_{\theta_i} l(\theta, x) \partial_{\theta_j} l(\theta, x)]. \quad (2.24)$$

**Proposición 2.3.4.** La matriz de información de Fisher (2.24) también se puede escribir como:

$$\mathcal{I}_{ij}(\theta) = -E_{\theta} [\partial_{\theta_j} \partial_{\theta_i} l(\theta, x)]. \quad (2.25)$$

*Demostración.*

$$\begin{aligned} -E_{\theta} [\partial_{\theta_j} \partial_{\theta_i} l(\theta, x)] &= -\int_{\Omega} p_{\theta}(x) \partial_{\theta_j} \partial_{\theta_i} \ln(p_{\theta}(x)) dx \\ &= -\int_{\Omega} p_{\theta}(x) \partial_{\theta_j} \left( \frac{1}{p_{\theta}(x)} \partial_{\theta_i} p_{\theta}(x) \right) dx \\ &= -\int_{\Omega} p_{\theta}(x) \left( \frac{p_{\theta}(x) \partial_{\theta_j} \partial_{\theta_i} p_{\theta}(x) - \partial_{\theta_j} p_{\theta}(x) \partial_{\theta_i} p_{\theta}(x)}{[p_{\theta}(x)]^2} \right) dx \\ &= -\int_{\Omega} p_{\theta}(x) \left( \frac{\partial_{\theta_j} \partial_{\theta_i} p_{\theta}(x)}{p_{\theta}(x)} - \frac{\partial_{\theta_j} p_{\theta}(x)}{p_{\theta}(x)} \frac{\partial_{\theta_i} p_{\theta}(x)}{p_{\theta}(x)} \right) dx \\ &= -\int_{\Omega} p_{\theta}(x) \left( \frac{\partial_{\theta_j} \partial_{\theta_i} p_{\theta}(x)}{p_{\theta}(x)} - \partial_{\theta_j} l(\theta, x) \partial_{\theta_i} l(\theta, x) \right) dx \\ &= -\int_{\Omega} (\partial_{\theta_j} \partial_{\theta_i} p_{\theta}(x) - p_{\theta}(x) \partial_{\theta_j} l(\theta, x) \partial_{\theta_i} l(\theta, x)) dx \\ &= -\partial_{\theta_j} \partial_{\theta_i} \left( \int_{\Omega} p_{\theta}(x) dx \right) + \int_{\Omega} p_{\theta}(x) \partial_{\theta_i} l(\theta, x) \partial_{\theta_j} l(\theta, x) dx \\ &= E_{\theta} [\partial_{\theta_i} l(\theta, x) \partial_{\theta_j} l(\theta, x)], \end{aligned}$$

donde en la última igualdad se utilizó que  $\int_{\Omega} p_{\theta}(x) dx = 1$ . ■

Ahora veremos la estructura geométrica inducida por la entropía relativa.

**Proposición 2.3.5.** Sobre  $M$  se tiene una estructura de variedad estadística dada por:

$${}^D g_{ij}(\theta) = E_{\theta} [\partial_{\theta_i} l(\theta, x) \partial_{\theta_j} l(\theta, x)], \quad (2.26)$$

$${}^D \Gamma_{ij,k}(\theta) = E_{\theta} [\partial_{\theta_k} l(\theta, x) (\partial_{\theta_i} \partial_{\theta_j} l(\theta, x) + \partial_{\theta_j} l(\theta, x) \partial_{\theta_i} l(\theta, x))], \quad (2.27)$$

$${}^D \Gamma_{ij,k}^*(\theta) = E_{\theta} [\partial_{\theta_j} \partial_{\theta_i} l(\theta, x) \partial_{\theta_k} l(\theta, x)]. \quad (2.28)$$

*Demostración.* Como ya sabemos que (2.23) es una función de divergencia entonces por el Teorema 2.2.2 esta induce una métrica Riemanniana con conexiones conjugadas. Por (c) de lo anterior tenemos que en estas coordenadas las componentes de la métrica están dadas por  ${}^D g_{ij}(\theta) = E_{\theta} [\partial_{\theta_i} l(\theta, x) \partial_{\theta_j} l(\theta, x)]$ . Veamos como quedan las conexiones conjugadas. Usando la primera igualdad obtenida en (c) se tiene,

$${}^D \Gamma_{ij,k}(\theta) = \partial_{\theta_i} (-\partial_{\theta_j} \partial_{\theta_k} D(\theta, \eta)) \Big|_{\theta=\eta} = - \int_{\Omega} \partial_{\theta_i} \partial_{\theta_j} p_{\theta}(x) \partial_{\theta_k} l(\eta, x) dx \Big|_{\theta=\eta}. \quad (2.29)$$

Note que,

$$\begin{aligned} \partial_{\theta_i} \partial_{\theta_j} l(\theta, x) &= \partial_{\theta_i} \left( \frac{\partial_{\theta_j} p_{\theta}(x)}{p_{\theta}(x)} \right) \\ &= \frac{p_{\theta}(x) \partial_{\theta_i} \partial_{\theta_j} p_{\theta}(x) - \partial_{\theta_j} p_{\theta}(x) \partial_{\theta_i} p_{\theta}(x)}{[p_{\theta}(x)]^2} \\ &= \frac{\partial_{\theta_i} \partial_{\theta_j} p_{\theta}(x)}{p_{\theta}(x)} - \partial_{\theta_j} l(\theta, x) \partial_{\theta_i} l(\theta, x), \end{aligned}$$

de aquí se tiene que

$$\partial_{\theta_i} \partial_{\theta_j} p_{\theta}(x) = p_{\theta}(x) \left( \partial_{\theta_i} \partial_{\theta_j} l(\theta, x) + \partial_{\theta_j} l(\theta, x) \partial_{\theta_i} l(\theta, x) \right).$$

Luego reemplazando en (2.29) se tiene que

$${}^D\Gamma_{ij,k}(\theta) = E_{\theta} \left[ \partial_{\theta_k} l(\theta, x) \left( \partial_{\theta_i} \partial_{\theta_j} l(\theta, x) + \partial_{\theta_j} l(\theta, x) \partial_{\theta_i} l(\theta, x) \right) \right].$$

Por otro lado

$${}^D\Gamma_{ij,k}^*(\theta) = \partial_{\eta_j} \left( -\partial_{\theta_k} \partial_{\eta_i} D(\theta, \eta) \right) \Big|_{\theta=\eta} = \int_{\Omega} \partial_{\eta_j} \partial_{\eta_i} \ln(p_{\eta}(x)) \partial_{\theta_k} p_{\theta}(x) dx \Big|_{\theta=\eta},$$

y por tanto,

$${}^D\Gamma_{ij,k}^*(\theta) = E_{\theta} \left[ \partial_{\theta_j} \partial_{\theta_i} l(\theta, x) \partial_{\theta_k} l(\theta, x) \right].$$

Y con esto tenemos definida totalmente la geometría sobre la variedad  $M$ . ■

Considere la densidad  $p_{\theta}(x)$  y sea  $y = f(x)$  la transformación de la variable aleatoria  $x$  a la variable aleatoria  $y$ , dicha transformación en general no es inyectiva, y por tanto no siempre es posible recuperar  $x$  a partir de  $y$ . Si no es inyectiva se dice que se pierde la información por la función  $f$ . Teniendo en cuenta esto introducimos la siguiente definición.

**Definición 2.3.6.** Una función  $y = f(x)$  se llama **estadística suficiente** cuando la densidad de probabilidad  $p_{\theta}(x)$  se puede descomponer como  $p_{\theta}(x) = \tilde{p}_{\theta}(y)g(x)$ . Es decir, la información que se pierde al pasar de  $x$  a  $y$  no afecta al parámetro  $\theta$ . En este sentido  $y$  es suficiente para obtener a  $\theta$ .

Ahora sea  $\tilde{p}_{\theta}(y)$  la densidad de probabilidad para  $y$ . Así la divergencia de  $D(\theta, \eta)$  entre  $p_{\theta}(x)$  y  $p_{\eta}(x)$  cambia a  $\tilde{D}(\theta, \eta)$  que es entre  $\tilde{p}_{\theta}(y)$  y  $\tilde{p}_{\eta}(y)$ . Si no hay pérdida de información entonces se esperaría que la divergencia entre las dos densidades fuera la misma, cuando hay pérdida se esperaría que sería más difícil distinguir a  $\theta$  de  $\eta$ , lo cual significa que

$$\tilde{D}(\theta, \eta) \leq D(\theta, \eta). \quad (2.30)$$

La ecuación (2.30) se conoce como **monotonidad de la información**. Además, la igualdad se da cuando la transformación  $f$  es una estadística suficiente.

La métrica de Fisher tiene muchas derivaciones y en algún sentido es la única (salvo por un factor constante) métrica Riemanniana sobre la variedad de información que es compatible con la distribución de probabilidad  $p_{\theta}$ . La prueba del siguiente teorema puede consultarse en [A16].

**Teorema 2.3.7.** *Cualquier divergencia que satisface la monotonicidad de la información induce la métrica de Fisher salvo por un factor constante.*

Ahora veremos dos ejemplos usuales de esta métrica.

**Ejemplo 2.3.8. Familia de distribuciones exponenciales:** Es una familia estadística parametrizada por  $\theta \in \Theta$ , con  $\Theta \subset \mathbb{R}^n$  abierto. Las distribuciones están dadas por:

$$\mathcal{E} = \left\{ p_{\theta}(x) = e^{h(x) \cdot \theta + k(x) - \psi(\theta)} \quad \text{para } \theta \in \Theta \right\}, \quad (2.31)$$

con  $x$  una variable aleatoria,  $h(x) \cdot \theta = \sum_{i=1}^N h_i(x) \theta_i$  donde las funciones  $\{h_i\}_{i=1}^N$  son linealmente independientes en algún soporte en común y  $k$  es una función de  $x$ . Además  $\psi$  es

un factor de normalización con el cual se tiene  $\int_{\Omega} p_{\theta}(x) dx = 1$  para todo  $\theta$ , por tanto, se tiene que

$$\psi(\theta) = \ln \left( \int_{\Omega} e^{h(x) \cdot \theta + k(x)} dx \right),$$

Veamos que  $\psi$  es una función convexa. Como  $1 = \int_{\Omega} p_{\theta}(x) dx$ , entonces derivando respecto a  $\theta_i$  se tiene

$$\begin{aligned} 0 &= \int_{\Omega} \left( h_i(x) - \frac{\partial \psi(\theta)}{\partial \theta_i} \right) p_{\theta}(x) dx \\ &= \int_{\Omega} h_i(x) p_{\theta}(x) dx - \frac{\partial \psi(\theta)}{\partial \theta_i} \int_{\Omega} p_{\theta}(x) dx \\ &= E_{\theta}[h_i(x)] - \frac{\partial \psi(\theta)}{\partial \theta_i}, \end{aligned}$$

por tanto, se tiene que  $E_{\theta}[h_i(x)] = \frac{\partial \psi(\theta)}{\partial \theta_i}$ . Derivando la expresión  $\int_{\Omega} \left( h_i(x) - \frac{\partial \psi(\theta)}{\partial \theta_i} \right) p_{\theta}(x) dx$  respecto a  $\theta_j$  se tiene:

$$\begin{aligned} 0 &= \int_{\Omega} \left( -\frac{\partial^2 \psi(\theta)}{\partial \theta_j \partial \theta_i} p_{\theta}(x) + \left( h_i(x) - \frac{\partial \psi(\theta)}{\partial \theta_i} \right) \frac{\partial p_{\theta}(x)}{\partial \theta_j} \right) dx \\ &= -\frac{\partial^2 \psi(\theta)}{\partial \theta_j \partial \theta_i} + \int_{\Omega} \left( h_i(x) - \frac{\partial \psi(\theta)}{\partial \theta_i} \right) \left( h_j(x) - \frac{\partial \psi(\theta)}{\partial \theta_j} \right) p_{\theta}(x) dx \\ &= -\frac{\partial^2 \psi(\theta)}{\partial \theta_j \partial \theta_i} + \int_{\Omega} (h_i(x) - E_{\theta}[h_i(x)]) (h_j(x) - E_{\theta}[h_j(x)]) p_{\theta}(x) dx \\ &= -\frac{\partial^2 \psi(\theta)}{\partial \theta_j \partial \theta_i} + \text{Var}[h_i(x)], \end{aligned}$$

así  $\frac{\partial^2 \psi(\theta)}{\partial \theta_j \partial \theta_i} = \text{Var}[h_i(x)]$  donde  $\text{Var}$  es la matriz de covarianza la cual sabemos que siempre es definida positiva y por tanto  $\psi$  es convexa. Por otro lado, note que:

$$\frac{\partial(\ln(p_{\theta}(x)))}{\partial \theta_i} = h_i(x) - \frac{\partial \psi(\theta)}{\partial \theta_i} = h_i(x) - E_{\theta}[h_i(x)],$$

reemplazando en (2.26) se tiene

$$\mathcal{I}_{\mathcal{E}}(\theta) = \int_{\Omega} (h_i(x) - E_{\theta}[h_i(x)]) (h_j(x) - E_{\theta}[h_j(x)]) p_{\theta}(x) dx = \frac{\partial^2 \psi(\theta)}{\partial \theta_j \partial \theta_i}.$$

Por tanto, se obtiene una variedad Hessiana.

**Ejemplo 2.3.9. Familia de distribuciones Gaussianas:** Ahora veremos un ejemplo muy bonito por la métrica que se obtiene. Considere la función de densidad de probabilidad Gaussiana para una variable aleatoria  $x$  dada por:

$$p_{(\mu, \sigma)}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (2.32)$$

El conjunto de todas las distribuciones Gaussianas forma una variedad 2-dimensional donde los puntos son las funciones de densidad (2.32) y  $(\mu, \sigma)$  es un sistema de coordenadas para la variedad, donde  $\sigma > 0$ . Así este espacio es homeomorfo al semi plano

superior  $\Theta = \{(\mu, \sigma) \in \mathbb{R}^2 : \sigma > 0\}$ . Note que  $l((\mu, \sigma), x) = -\frac{1}{2} \ln(2\pi) - \ln(\sigma) - \frac{(x-\mu)^2}{2\sigma^2}$ . Para calcular la métrica usaremos (2.25), luego necesitamos calcular las segundas derivadas de  $l$

$$\begin{aligned}\partial_\mu^2 l((\mu, \sigma), x) &= -\frac{1}{\sigma^2}, \\ \partial_\mu \partial_\sigma l((\mu, \sigma), x) &= -2 \frac{x - \mu}{\sigma^3}, \\ \partial_\sigma \partial_\mu l((\mu, \sigma), x) &= -2 \frac{x - \mu}{\sigma^3}, \\ \partial_\sigma^2 l((\mu, \sigma), x) &= \frac{1}{\sigma^2} - 3 \frac{(x - \mu)^2}{\sigma^4}.\end{aligned}$$

Colocando estas expresiones en (2.25) se obtiene que la matriz de Fisher es:

$$\mathcal{I}_\Theta((\mu, \sigma)) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix},$$

la cual se corresponde a la métrica de Poincaré para el semi plano superior, que se escribe como:

$$(ds)^2 = \frac{(d\mu)^2 + 2(d\sigma)^2}{\sigma^2}.$$

## CAPÍTULO 3

# ALGORITMOS DE OPTIMIZACIÓN GEOMÉTRICA

### 3.1. Introducción a la Optimización

Considere  $M$  una variedad  $n$ -dimensional y  $f : M \rightarrow \mathbb{R}$ . El problema estándar de optimización es encontrar el valor donde se obtiene el mínimo (o máximo) de una función, esto es, encontrar

$$q^* \in \arg \min_{q \in M} f(q). \quad (3.1)$$

Los algoritmos de optimización son operaciones dadas paso por paso las cuales terminan después de cierto número de pasos y que permite generar una solución rápida con cierta exactitud y precisión del problema anterior.

A continuación, vamos a describir brevemente algunos de los métodos de optimización más usados, tomando las ideas de [HLW06], [SMDH13] y [L19].

#### 3.1.1. Métodos con el gradiente

En la literatura, uno de los métodos más populares para solucionar el problema de minimizar una función, es el **método del descenso del gradiente** donde el punto máximo o mínimo puede ser local o global. El valor mínimo de una función se obtiene rápidamente si se mueve en la dirección negativa del gradiente y viceversa, el máximo se obtiene al moverse en la dirección del gradiente. Sea  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , con  $f \in C^1$  podemos iniciar en un punto  $x_0$  y empezamos a iterar mediante la siguiente relación,

$$x_{k+1} = x_k - \tau \nabla f(x_k), \quad (3.2)$$

donde el término  $\tau$  es el tamaño de paso el cual es menor a 1.

Ahora consideremos un resultado de geometría Riemanniana:

**Proposición 3.1.1.** *Sea  $(M, g)$  una variedad Riemanniana. Sea  $f : M \rightarrow \mathbb{R}$  una función sobre la variedad entonces el gradiente en  $M$  viene dado por*

$$\text{Grad}(f) = g^{-1} \nabla f$$

con  $\nabla f$  el gradiente en  $\mathbb{R}^n$ .

Veamos una generalización inmediata del método del gradiente que nos da un algoritmo para minimizar funciones definidas sobre variedades Riemannianas. Considere  $M$  una variedad estadística parametrizada por  $\theta = (\theta_1, \dots, \theta_n)$ . Note que por lo mencionado en la Sección 2.3 las variedades estadísticas están equipadas naturalmente con la métrica de Fisher, esto es particularmente relevante en el caso de optimizar funciones sobre variedades estadísticas. El método que vamos a introducir va a ser importante como método para obtener los parámetros que minimizan el negativo de la esperanza de la función log-verosimilitud  $L(\theta) = -E_\theta[l(\theta, x)]$ . La función  $L$  es una función escalar sobre  $M$  y usualmente es suave. Por tanto, a partir de lo discutido anteriormente y la Proposición 3.1.1 es natural dar la siguiente definición:

**Definición 3.1.2.** *Sea  $M$  una variedad  $n$ -dimensional. El **gradiente natural** está dado por  $\nabla_N L = \mathcal{I}^{-1} \nabla L$  donde  $\nabla$  es el gradiente usual en  $\mathbb{R}^n$  e  $\mathcal{I}$  es la matriz de información de Fisher.*

De esta forma podemos considerar el **método del descenso por el gradiente natural** dado por la expresión:

$$\theta_{k+1} = \theta_k - \tau \mathcal{I}^{-1} \nabla L(\theta_k). \quad (3.3)$$

El método del descenso por el gradiente natural tiene la ventaja de que no se va a atorar en los puntos donde el gradiente normal lo hace, es decir, cuando el gradiente de  $f$  es casi 0. Su desventaja es que hay que calcular la matriz de información de Fisher en cada iteración, y normalmente eso involucra calcular Hessiana (como en el Ejemplo 2.3.8) y para problemas en muchas dimensiones esto es muy complicado de hacer. Es por esto que a pesar de que el gradiente natural es muy bueno a nivel teórico, a nivel práctico se prefieren los algoritmos que sólo involucran información sobre el gradiente Euclideo de  $f$ .

### 3.1.2. Métodos con el gradiente más momentum

Primero consideramos el método del **momento clásico** el cual introduce un factor de velocidad que hace que la rapidez de convergencia del método del gradiente se acelere. El método está dado por:

$$p_{k+1} = \mu p_k - \tau \nabla f(q_k), \quad (3.4)$$

$$q_{k+1} = q_k + \tau p_{k+1}, \quad (3.5)$$

donde  $\tau > 0$  es el tamaño de paso,  $p_k$  el momento y  $\mu \in (0, 1)$  se llama *coeficiente de disipación*.

Otro método es el **método de gradiente acelerado de Nesterov** el cual es una versión modificada del método de momento y está dado por:

$$p_{k+1} = \mu_k p_k - \tau \nabla f(q_k + \mu_k p_k), \quad (3.6)$$

$$q_{k+1} = q_k + \tau p_{k+1}. \quad (3.7)$$

Lo primero a notar es que el método de Nesterov evalúa el gradiente en un punto “futuro” ( $q_k + \mu_k p_k$ ), mientras que el método de momento lo evalúa en el punto “actual” ( $q_k$ ), es decir, estamos usando un punto cercano a  $q_k$  que se puede ver como una anticipación que acelera el proceso de convergencia. A continuación, se puede ver de manera esquemática cómo se comporta los dos métodos mencionados anteriormente.



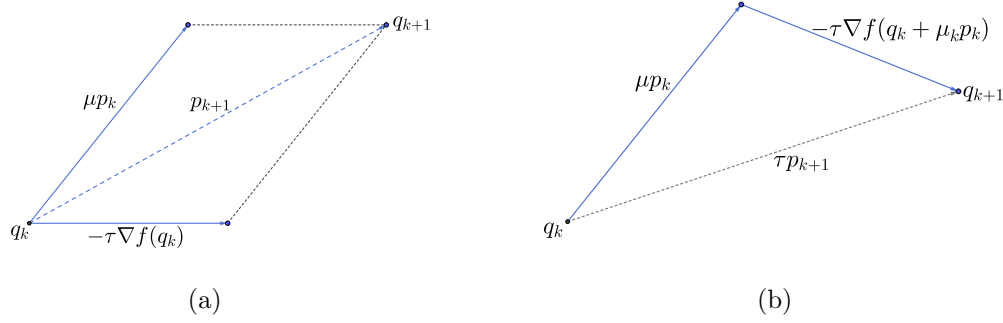


Figura 3.1: (a) Método de momento clásico. (b) Método del gradiente acelerado de Nesterov.

### 3.1.3. De los algoritmos de optimización a los sistemas dinámicos

A continuación, para cada uno de los métodos mencionados anteriormente encontraremos una ecuación diferencial que da su límite continuo y donde las ideas son tomadas de [SBC16].

**Proposición 3.1.3.** *El límite continuo del método del gradiente (3.2) está dado por:*

$$\dot{x}(t) + \nabla f(x(t)) = 0. \quad (3.8)$$

*Demostración.* Considere el ansatz  $x_k \approx x(k\tau) = x(t)$ , luego  $x_{k+1} \approx x((k+1)\tau) = x(t + \tau)$ . Realizando la expansión de Taylor se tiene:

$$x_{k+1} - x_k \approx x(t + \tau) - x(t) = \dot{x}(t)\tau + \frac{1}{2}\ddot{x}(t)\tau^2 + \mathcal{O}(\tau^3), \quad (3.9)$$

reemplazando en (3.2) se obtiene

$$\dot{x}(t)\tau + \frac{1}{2}\ddot{x}(t)\tau^2 + \mathcal{O}(\tau^3) = -\tau \nabla f(x(t)),$$

de aquí se tiene,

$$(\dot{x}(t) + \nabla f(x(t)))\tau + \frac{1}{2}\ddot{x}(t)\tau^2 + \mathcal{O}(\tau^3) = 0,$$

despreciando los términos de orden superior se tiene  $(\dot{x}(t) + \nabla f(x(t)))\tau = 0$  y de acá se tiene  $\dot{x}(t) + \nabla f(x(t)) = 0$ . ■

Si realizamos este mismo procedimiento para el gradiente natural (3.3) se obtiene:

$$\dot{\theta}(t) + \mathcal{I}^{-1}\nabla L(\theta(t)) = 0. \quad (3.10)$$

**Proposición 3.1.4.** *El método del momento dado por las ecuaciones (3.4) y (3.5), tiene como límite continuo la ecuación diferencial*

$$\ddot{q} + \gamma \dot{q} + \nabla f(q) = 0, \quad (3.11)$$

donde  $\gamma > 0$  es una constante que está relacionada con  $\mu$  mediante la expresión  $\mu = e^{-\gamma\tau}$ .

*Demostración.* Multiplicando la ecuación (3.4) por  $\tau$  y usando (3.5) se tiene

$$q_{k+1} - q_k = \tau \mu p_k - \tau^2 \nabla f(q_k),$$

restando a ambos lados  $\tau p_k$

$$q_{k+1} - q_k - \tau p_k = \tau(\mu - 1)p_k - \tau^2 \nabla f(q_k),$$

usando nuevamente (3.5) se obtiene,

$$q_{k+1} - q_k - (q_k - q_{k-1}) = (\mu - 1)(q_k - q_{k-1}) - \tau^2 \nabla f(q_k). \quad (3.12)$$

Además, considere el ansatz  $q_k \approx q(k\tau) = q(t)$  y de acá se tiene  $q_{k+1} \approx q((k+1)\tau) = q(t+\tau)$  y de la misma forma  $q_{k-1} \approx q(t-\tau)$ . Expandiendo en la serie de Taylor se tiene la ecuación (3.9) y además,

$$q_k - q_{k-1} \approx q(t) - q(t-\tau) = \dot{q}(t)\tau - \frac{1}{2}\ddot{q}(t)\tau^2 + \mathcal{O}(\tau^3),$$

reemplazando en (3.12) se obtiene:

$$\dot{q}\tau + \frac{\ddot{q}\tau^2}{2} + \mathcal{O}(\tau^3) - \left( \dot{q}\tau - \frac{\ddot{q}\tau^2}{2} + \mathcal{O}(\tau^3) \right) = (\mu - 1) \left( \dot{q}\tau - \frac{\ddot{q}\tau^2}{2} + \mathcal{O}(\tau^3) \right) - \tau^2 \nabla f(q),$$

acomodando términos se tiene,

$$\ddot{q}\tau^2 + \mathcal{O}(\tau^3) = (\mu - 1)\dot{q}\tau - (\mu - 1)\frac{\ddot{q}\tau^2}{2} + \mathcal{O}(\tau^3) - \tau^2 \nabla f(q).$$

Note que como  $\mu \in (0, 1)$  entonces se puede escribir como  $\mu = e^{-\gamma\tau}$ , que al hacer su expansión en Taylor queda  $\mu - 1 = -\gamma\tau + \mathcal{O}(\tau^2)$ , luego reemplazando en la expresión anterior se tiene,

$$\ddot{q}\tau^2 + \mathcal{O}(\tau^3) = \gamma\dot{q}\tau^2 + \mathcal{O}(\tau^3) - \tau^2 \nabla f(q),$$

agrupando términos se tiene

$$(\ddot{q} + \gamma\dot{q} + \nabla f(q))\tau^2 + \mathcal{O}(\tau^3) = 0,$$

despreciando los términos de orden superior se tiene el resultado. ■

Note que al tomar  $\mu = e^{-\gamma\tau}$  estamos dando sentido al llamar  $\mu$  coeficiente de disipación ya que en la ecuación (3.11), físicamente  $\gamma$  está asociado a la fricción del sistema. Es más, veamos que no se conserva la energía. Considere la energía del sistema  $H = \frac{p^2}{2} + f(q)$  y tomando  $p = \dot{q}$  la ecuación (3.11) se puede reescribir como  $\dot{p} = -\gamma p - \nabla f(q)$ , entonces

$$\dot{H} = p\dot{p} + \nabla f(q)\dot{q} = p(-\gamma p - \nabla f(q)) + \nabla f(q)p = -\gamma p^2. \quad (3.13)$$

**Proposición 3.1.5.** *El método de gradiente acelerado de Nesterov dado por las ecuaciones (3.6) y (3.7) tiene como límite continuo la ecuación diferencial*

$$\ddot{q} + \frac{3}{t}\dot{q} + \nabla f(q) = 0. \quad (3.14)$$

*Demostración.* Tome  $\mu_k = \frac{k-1}{k+2}$  y reacomodando se puede reescribir el método de Nesterov como:

$$q_k = p_{k-1} - \tau \nabla f(p_{k-1}), \quad (3.15)$$

$$p_k = q_k + \frac{k-1}{k+2}(q_k - q_{k-1}). \quad (3.16)$$

Ahora cambiando los índices en (3.15) por  $k+1$  y reemplazando (3.16) en (3.15) se tiene

$$q_{k+1} = q_k + \frac{k-1}{k+2}(q_k - q_{k-1}) - \tau \nabla f(p_k).$$

Dividiendo por  $\sqrt{\tau}$  se tiene:

$$\frac{q_{k+1} - q_k}{\sqrt{\tau}} = \frac{k-1}{k+2} \frac{q_k - q_{k-1}}{\sqrt{\tau}} - \sqrt{\tau} \nabla f(p_k). \quad (3.17)$$

Ahora sea  $q(t)$  una curva suave definida para  $t \geq 0$ . Tome  $t = k\sqrt{\tau}$  entonces note que

$$\frac{k-1}{k+2} = \frac{\frac{t}{\sqrt{\tau}} - 1}{\frac{t}{\sqrt{\tau}} + 2} = \frac{t - \sqrt{\tau}}{t + 2\sqrt{\tau}} = \frac{t + 2\sqrt{\tau} - 3\sqrt{\tau}}{t + 2\sqrt{\tau}} = 1 - \frac{3\sqrt{\tau}}{t + 2\sqrt{\tau}}. \quad (3.18)$$

Además considere el ansatz  $q_k \approx q(k\sqrt{\tau}) = q(t)$  y de acá se tiene  $q_{k+1} \approx q((k+1)\sqrt{\tau}) = q(t + \sqrt{\tau})$  y de la misma forma  $q_{k-1} \approx q(t - \sqrt{\tau})$ . Expandiendo en la serie de Taylor se tiene

$$q_{k+1} - q_k \approx q(t + \sqrt{\tau}) - q(t) = \dot{q}(t)\sqrt{\tau} + \frac{1}{2}\ddot{q}(t)\tau + \mathcal{O}((\sqrt{\tau})^3),$$

$$q_k - q_{k-1} \approx q(t) - q(t - \sqrt{\tau}) = \dot{q}(t)\sqrt{\tau} - \frac{1}{2}\ddot{q}(t)\tau + \mathcal{O}((\sqrt{\tau})^3),$$

dividiendo por  $\sqrt{\tau}$  en estas expresiones se obtiene:

$$\frac{q_{k+1} - q_k}{\sqrt{\tau}} = \dot{q}(t) + \frac{1}{2}\ddot{q}(t)\sqrt{\tau} + \mathcal{O}(\sqrt{\tau}), \quad (3.19)$$

$$\frac{q_k - q_{k-1}}{\sqrt{\tau}} = \dot{q}(t) - \frac{1}{2}\ddot{q}(t)\sqrt{\tau} + \mathcal{O}(\sqrt{\tau}). \quad (3.20)$$

Ahora para el término  $\nabla f(p_k)$  utilizamos (3.16), (3.18) y (3.20) y expandiendo en su serie de Taylor se obtiene:

$$\sqrt{\tau} \nabla f(p_k) = \sqrt{\tau} \nabla f \left( q_k + \sqrt{\tau} \left( 1 - \frac{3\sqrt{\tau}}{t + 2\sqrt{\tau}} \right) \frac{q_k - q_{k-1}}{\sqrt{\tau}} \right) = \sqrt{\tau} \nabla f(q(t)) + \mathcal{O}(\sqrt{\tau}),$$

reemplazando todo esto en (3.17) se tiene:

$$\dot{q} + \frac{\sqrt{\tau}}{2}\ddot{q} + \mathcal{O}(\sqrt{\tau}) = \left( 1 - \frac{3\sqrt{\tau}}{t + 2\sqrt{\tau}} \right) \left( \dot{q} - \frac{\sqrt{\tau}}{2}\ddot{q} + \mathcal{O}(\sqrt{\tau}) \right) - \sqrt{\tau} \nabla f(q) + \mathcal{O}(\sqrt{\tau}),$$

acomodando términos se obtiene:

$$\left( \ddot{q} + \frac{3}{t + 2\sqrt{\tau}}\dot{q} + \nabla f(q) \right) \sqrt{\tau} + \mathcal{O}(\sqrt{\tau}) = 0,$$

despreciando los términos de orden superior se tiene  $\ddot{q} + \frac{3}{t+2\sqrt{\tau}}\dot{q} + \nabla f(q) = 0$  y tomando límite cuando  $\tau \rightarrow 0$  se tiene el resultado. ■

Veamos que no se conserva la energía. Considere la energía del sistema  $H = \frac{p^2}{2} + f(q)$  y tomando  $p = \dot{q}$  la ecuación (3.14) se puede reescribir como  $\dot{p} = -\frac{3}{t}p - \nabla f(q)$ , entonces

$$\dot{H} = p\dot{p} + \nabla f(q)\dot{q} = p\left(-\frac{3}{t}p - \nabla f(q)\right) + \nabla f(q)p = -\frac{3}{t}p^2. \quad (3.21)$$

## 3.2. Optimización Hamiltoniana

En la sección anterior encontramos que los métodos de momento clásico y de Nestorov vistos, tienen como límite continuo ecuaciones diferenciales de segundo orden que disipan una energía (ver ecuaciones (3.13) y (3.21)), por tanto, es natural verlos desde una perspectiva Hamiltoniana, luego consideraremos los siguientes 3 casos.

### 3.2.1. Hamiltonianos dependientes del tiempo

Una cantidad importante en mecánica es la energía y como vimos en la sección anterior la función  $f$  juega el rol de energía potencial. Como necesitamos encontrar una energía cinética una propuesta interesante está dada por la función divergencia de Bregman (ver Ejemplo 2.2.3). Consideremos  $U \subset \mathbb{R}^n$  un dominio y sea  $h : U \rightarrow \mathbb{R}$  una función convexa. Considere la divergencia de Bregman  $B_h$  definida como en (2.11). El **Hamiltoniano de Bregman** tiene la forma (ver [WWJ16]):

$$H_{\alpha,\beta,\gamma} = e^{\alpha+\gamma} B_{h^*} (e^{-\gamma} p + \nabla h(q), \nabla h(q)) + e^{\alpha+\gamma+\beta} f(q), \quad (3.22)$$

donde  $(q, p) \in T^*U$ ,  $h^*$  esta dada por (2.14) y  $\alpha, \beta$  y  $\gamma$  son funciones que dependen del tiempo y además deben cumplir las condiciones:

$$\dot{\beta} \leq e^\alpha, \quad (3.23)$$

$$\dot{\gamma} = e^\alpha. \quad (3.24)$$

Estas condiciones tendrán sentido más adelante. La dinámica del Hamiltoniano de Bregman se obtiene de solucionar las ecuaciones (1.12). Recordando que  $q = \nabla h^*(\nabla h(q))$  se tiene:

$$\dot{q} = e^\alpha (\nabla h^* (\nabla h(q) + e^{-\gamma} p) - q), \quad (3.25)$$

$$\dot{p} = -e^{\alpha+\gamma} \nabla^2 h(q) (\nabla h^* (\nabla h(q) + e^{-\gamma} p) - q) + e^\alpha p - e^{\alpha+\beta+\gamma} \nabla f(q). \quad (3.26)$$

Reacomodando la ecuación (3.25) se tiene:

$$e^{-\alpha} \dot{q} + q = \nabla h^* (\nabla h(q) + e^{-\gamma} p), \quad (3.27)$$

usando el hecho de que  $p = \nabla h(\nabla h^*(p))$  se tiene:

$$\nabla h (e^{-\alpha} \dot{q} + q) = \nabla h(q) + e^{-\gamma} p, \quad (3.28)$$

usando esta expresión y derivando respecto al tiempo se tiene:

$$\frac{d}{dt} (\nabla h (e^{-\alpha} \dot{q} + q)) = \frac{d}{dt} (\nabla h(q) + e^{-\gamma} p) = \nabla^2 h(q) \dot{q} - \dot{\gamma} e^{-\gamma} p + e^{-\gamma} \dot{p}. \quad (3.29)$$

Ahora si reemplazamos (3.27) en (3.26) se tiene

$$\dot{p} = -e^\gamma \nabla^2 h(q) \dot{q} + e^\alpha p - e^{\alpha+\beta+\gamma} \nabla f(q). \quad (3.30)$$

Ahora realizando la derivada de la izquierda de (3.29) se tiene

$$\nabla^2 h (e^{-\alpha} \dot{q} + q) (\dot{q} + e^{-\alpha} \ddot{q} - \dot{\alpha} e^{-\alpha} \dot{q}) = \nabla^2 h(q) \dot{q} - \dot{\gamma} e^{-\gamma} p + e^{-\gamma} \dot{p},$$

reacomodando términos y reemplazando (3.30) en la expresión anterior se tiene:

$$e^{-\alpha} \nabla^2 h (e^{-\alpha} \dot{q} + q) (\ddot{q} + (e^\alpha - \dot{\alpha}) \dot{q}) = (e^\alpha - \dot{\gamma}) e^{-\gamma} p - e^{\alpha+\beta} \nabla f(q).$$

Como  $h$  es convexa se tiene que la matriz  $\nabla^2 h (e^{-\alpha} \dot{q} + q)$  es invertible y despejando  $e^{-\gamma} p$  de (3.28) se obtiene

$$\begin{aligned} & \ddot{q} + (e^\alpha - \dot{\alpha}) \dot{q} + e^{2\alpha+\beta} [\nabla^2 h (e^{-\alpha} \dot{q} + q)]^{-1} \nabla f(q) + \\ & e^\alpha (\dot{\gamma} - e^\alpha) [\nabla^2 h (e^{-\alpha} \dot{q} + q)]^{-1} [\nabla h (e^{-\alpha} \dot{q} + q) - \nabla h(q)] = 0. \end{aligned} \quad (3.31)$$

Esta es una ecuación diferencial de segundo orden obtenida de las ecuaciones de Hamilton. Note que partiendo de (3.29) y usando (3.24), (3.25), (3.26) se obtiene:

$$\frac{d}{dt} (\nabla h (e^{-\alpha} \dot{q} + q)) = -e^{\alpha+\beta} \nabla f(q). \quad (3.32)$$

**Teorema 3.2.1.** Sea  $q^*$  como (3.1), entonces la solución a la ecuación (3.32) satisface:

$$f(q(t)) - f(q^*) \leq \mathcal{O}(e^{-\beta(t)}).$$

*Demostración.* Considere la función de energía

$$\mathcal{E}_t(q) = B_h(q^*, q + e^{-\alpha} \dot{q}) + e^\beta (f(q) - f(q^*)). \quad (3.33)$$

Note que esta es una función de Lyapunov, pues  $\mathcal{E}_t(q^*) = 0$  y  $\mathcal{E}_t(q) > 0$  para todo  $q \in U - \{q^*\}$  (pues  $f$  y  $h$  son convexas). Además, note que (3.33) es casi el Hamiltoniano de Bregman (3.22). Derivando con respecto al tiempo y usando (3.32) se tiene:

$$\begin{aligned} \dot{\mathcal{E}}_t(q) &= - \left\langle q^* - q - e^{-\alpha} \dot{q}, \frac{d}{dt} (\nabla h (e^{-\alpha} \dot{q} + q)) \right\rangle + \dot{\beta} e^\beta (f(q) - f(q^*)) + e^\beta \langle \nabla f(q), \dot{q} \rangle \\ &= - \langle q^* - q - e^{-\alpha} \dot{q}, -e^{\alpha+\beta} \nabla f(q) \rangle + \dot{\beta} e^\beta (f(q) - f(q^*)) + e^\beta \langle \nabla f(q), \dot{q} \rangle \\ &= e^{\alpha+\beta} \langle q^* - q, \nabla f(q) \rangle + \dot{\beta} e^\beta (f(q) - f(q^*)) \\ &= -e^{\alpha+\beta} (f(q^*) - f(q) - \langle q^* - q, \nabla f(q) \rangle) + \dot{\beta} e^\beta (f(q) - f(q^*)) + e^{\alpha+\beta} (f(q^*) - f(q)) \\ &= -e^{\alpha+\beta} B_f(q^*, q) + (\dot{\beta} - e^\alpha) e^\beta (f(q) - f(q^*)). \end{aligned}$$

Usando la condición (3.23) y que  $f(q) - f(q^*) \geq 0$  (por ser  $q^*$  el punto donde  $f$  alcanza el mínimo), vemos que  $(\dot{\beta} - e^\alpha) e^\beta (f(q) - f(q^*)) \leq 0$ . Además como  $f$  es convexa entonces  $B_f(q^*, q) \geq 0$  y por tanto  $-e^{\alpha+\beta} B_f(q^*, q) \leq 0$ , luego se tiene que  $\dot{\mathcal{E}}_t(q) \leq 0$ , así  $\mathcal{E}_t$  es decreciente, por tanto para todo  $t \geq t_0$  se tiene  $\mathcal{E}_t \leq \mathcal{E}_{t_0}$ . Como  $B_h(q^*, q + e^{-\alpha} \dot{q}) \geq 0$  entonces por (3.33) se tiene  $e^\beta (f(q) - f(q^*)) \leq \mathcal{E}_t(q) \leq \mathcal{E}_{t_0}(q)$  y por tanto

$$f(q) - f(q^*) \leq e^{-\beta} \mathcal{E}_{t_0}(q) = \mathcal{O}(e^{-\beta}).$$

■

Note que en la demostración se utilizó una función de Lyapunov. En muchas pruebas de convergencia utilizan una función de Lyapunov que se halla pensando en energías de sistemas mecánicos, pues en este caso se tomó una energía cinética dada por la divergencia de Bregman y una energía potencial dada por la función  $f$  a minimizar. Por otro lado, este teorema me da una razón de convergencia asociada a la solución de (3.32). Además, para  $\alpha$  dado podemos determinar a  $\gamma$  usando (3.24) y al colocar igualdad en (3.23) se tiene una razón de convergencia de

$$\mathcal{O}(e^{-\beta}) = \mathcal{O} \left( e^{-\int_{t_0}^t e^{\alpha(s)} ds} \right). \quad (3.34)$$

### 3.2.2. Hamiltonianos conformalmente simplécticos

Ahora consideraremos ejemplos para integradores simplécticos conformes y nos guiaremos de [ESRV19] y el apéndice A.

**Ejemplo 3.2.2.** Podemos hacer una descomposición de (1.22) de tal forma que llamamos  $C(z) = \Omega \nabla H(z)$  y  $D(z) = -cDz$  con los flujos  $\Phi_t^C$  y  $\Phi_t^D$  respectivamente. Primero consideramos una aproximación numérica a la parte conservativa  $\dot{z} = \Omega \nabla H(z)$ . Para ello considere  $\widehat{\Phi}_\tau^C$  un integrador numérico que aproxima a  $\Phi_t^C$  para intervalos de tiempo pequeños  $[t, t + \tau]$ . El método simpléctico de Euler es:

$$\begin{aligned} q_{k+1} &= q_k + \tau \nabla_p H(q_k, p_{k+1}) \\ p_{k+1} &= p_k - \tau \nabla_q H(q_k, p_{k+1}). \end{aligned}$$

Y luego para la parte disipativa  $\dot{z} = -cDz$  el cual me da las ecuaciones  $\dot{q} = 0$  y  $\dot{p} = -cp$ . Considerando  $\widehat{\Phi}_\tau^D$  un integrador numérico que aproxima a  $\Phi_t^D$  para intervalos de tiempo pequeños  $[t, t + \tau]$  y aplicando el método simpléctico de Euler se tiene:

$$\begin{aligned} q_{k+1} &= q_k \\ p_{k+1} &= e^{-c\tau} p_k. \end{aligned}$$

Así  $\widehat{\Phi}_\tau = \widehat{\Phi}_\tau^C \circ \widehat{\Phi}_\tau^D$  es un integrador numérico para (1.22) dado por:

$$\begin{aligned} q_{k+1} &= q_k + \tau \nabla_p H(q_k, p_{k+1}) \\ p_{k+1} &= e^{-c\tau} p_k - \tau \nabla_q H(q_k, p_{k+1}). \end{aligned} \tag{3.35}$$

Si consideramos un Hamiltoniano de la forma  $H(q, p) = T(p) + f(q)$  donde  $T$  es la energía cinética y  $f$  es la energía potencial, entonces la expresión anterior se reduce a

$$\begin{aligned} q_{k+1} &= q_k + \tau \nabla T(p_{k+1}) \\ p_{k+1} &= e^{-c\tau} p_k - \tau \nabla f(q_k). \end{aligned} \tag{3.36}$$

**Teorema 3.2.3.** El integrador (3.35) es un integrador simpléctico conforme.

*Demostración.* Note que de la ecuación (3.35) se tiene:

$$dq_{k+1} = dq_k + \tau H_{pq}(q_k, p_{k+1}) dq_k + \tau H_{pp}(q_k, p_{k+1}) dp_{k+1} \tag{3.37}$$

$$dp_{k+1} = e^{-c\tau} dp_k - \tau H_{qq}(q_k, p_{k+1}) dq_k - \tau H_{qp}(q_k, p_{k+1}) dp_{k+1} \tag{3.38}$$

Donde la notación  $H_{qp}(q_k, p_{k+1}) = \frac{\partial^2 H}{\partial q \partial p}(q_k, p_{k+1})$ . Reacomodando se tiene:

$$\begin{aligned} dq_{k+1} - \tau H_{pp}(q_k, p_{k+1}) dp_{k+1} &= (I + \tau H_{pq}) dq_k \\ (I + \tau H_{qp}) dp_{k+1} &= e^{-c\tau} dp_k - \tau H_{qq} dq_k, \end{aligned}$$

donde las derivadas de  $H$  están evaluadas en  $(q_k, p_{k+1})$ . Ahora haciendo el producto cuña entre estas dos expresiones se tiene:

$$(I + \tau H_{qp}) dq_{k+1} \wedge dp_{k+1} = e^{-c\tau} (I + \tau H_{pq}) dq_k \wedge dp_k$$

Como  $I + \tau H_{qp}$  es invertible entonces se tiene  $dq_{k+1} \wedge dp_{k+1} = e^{-c\tau} dq_k \wedge dp_k$  y por tanto es un integrador conforme. ■

**Ejemplo 3.2.4.** Considere el Hamiltoniano

$$H(q, p) = \frac{\|p\|^2}{2m} + f(q),$$

donde  $m$  es la masa de una partícula sujeta a un potencial  $f$ . Usando (1.21) se tiene

$$\begin{cases} \dot{q} = \frac{p}{m} \\ \dot{p} = -\nabla f(q) - cp, \end{cases} \quad (3.39)$$

este sistema puede ser escrito como  $\ddot{q} + c\dot{q} + \frac{\nabla f(q)}{m} = 0$  que al tomar  $m = 1$  se obtiene (3.11) que es el límite continuo del método de momento clásico. El método del momento clásico es un integrador simpléctico conforme para este sistema (Teorema 3.2.3), pero el método de gradiente acelerado de Nesterov no lo es.

**Teorema 3.2.5.** El método del gradiente acelerado de Nesterov dado por (3.6) y (3.7) no es un integrador simpléctico conforme para el sistema Hamiltoniano (3.39).

*Demostración.* Considere la discretización de (3.39) dada por:

$$p_{k+1} = e^{-\gamma\tau} p_k - \tau \nabla f \left( q_k + e^{-\gamma\tau} \frac{\tau}{2m} p_k \right), \quad (3.40)$$

$$q_{k+1} = q_k + \frac{\tau}{2m} p_{k+1}. \quad (3.41)$$

Para simplificar cálculos llame  $\tilde{q}_k = q_k + e^{-\gamma\tau} \frac{\tau}{2m} p_k$  y note que si  $\mu = e^{-\gamma\tau}$  entonces (3.40) y (3.41) es el método de Nesterov dado en (3.6) y (3.7). Ahora obtenemos:

$$dp_{k+1} = e^{-\gamma\tau} dp_k - \tau \nabla^2 f(\tilde{q}_k) d\tilde{q}_k, \quad (3.42)$$

$$dq_{k+1} = dq_k + \frac{\tau}{2m} dp_{k+1}, \quad (3.43)$$

donde  $d\tilde{q}_k = dq_k + e^{-\gamma\tau} \frac{\tau}{2m} dp_k$ . Note que

$$\begin{aligned} dq_{k+1} \wedge dp_{k+1} &= \left( dq_k + \frac{\tau}{2m} dp_{k+1} \right) \wedge dp_{k+1} \\ &= dq_k \wedge dp_{k+1} \\ &= e^{-\gamma\tau} dq_k \wedge dp_k - \tau \nabla^2 f(\tilde{q}_k) dq_k \wedge d\tilde{q}_k \\ &= e^{-\gamma\tau} dq_k \wedge dp_k - \tau \nabla^2 f(\tilde{q}_k) dq_k \wedge \left( dq_k + e^{-\gamma\tau} \frac{\tau}{2m} dp_k \right) \\ &= e^{-\gamma\tau} dq_k \wedge dp_k - e^{-\gamma\tau} \frac{\tau^2}{2m} \nabla^2 f(\tilde{q}_k) dq_k \wedge dp_k \\ &= e^{-\gamma\tau} \left( 1 - \frac{\tau^2}{2m} \nabla^2 f(\tilde{q}_k) \right) dq_k \wedge dp_k. \end{aligned}$$

Como  $dq_{k+1} \wedge dp_{k+1} \neq e^{-\gamma\tau} dq_k \wedge dp_k$  entonces (3.40) y (3.41) no puede ser simpléctico conforme. ■

### 3.2.3. Hamiltonianos de contacto

Ahora veremos cómo los métodos anteriores se pueden obtener mediante un Hamiltoniano de contacto, y esto se logra escogiendo un Hamiltoniano apropiado. El resultado es tomado de [BDFB19].

**Proposición 3.2.6.** *Sea  $\mathcal{H}$  un Hamiltoniano de contacto, entonces*

- (a) Para  $\mathcal{H} = H(q, p, t)$ , es decir,  $\mathcal{H}$  no depende de  $S$ , se recupera la dinámica simpléctica obtenida con las ecuaciones (1.12).
- (b) Para  $\mathcal{H} = H(q, p, t) + cS$  se recupera la dinámica de los sistemas conformalmente simplécticos obtenida de las ecuaciones (1.21).
- (c) Para  $\mathcal{H} = \frac{1}{2}\|p\|^2 + e^{2\alpha+\beta}f(q) + (e^\alpha - \dot{\alpha})S$  se recupera la dinámica de Bregman.
- (d) Para  $\mathcal{H} = \frac{1}{2}\|p\|^2 + f(q) + \frac{3}{t}S$  se recupera el límite continuo del método de Nesterov (3.14).

*Demostración.* (a) Como tenemos un Hamiltoniano de contacto sabemos que su dinámica está dada por las ecuaciones (1.26), luego si  $\mathcal{H} = H(q, p, t)$  vemos que  $\frac{\partial \mathcal{H}}{\partial S} = 0$  y por tanto las dos primeras ecuaciones en (1.26) nos dan (1.12) y la tercera ecuación está desacoplada de las dos primeras.

(b) Nuevamente utilizando (1.26) para  $\mathcal{H} = H(q, p, t) + cS$  se tiene  $\frac{\partial \mathcal{H}}{\partial S} = c$  y por tanto las dos primeras ecuaciones de (1.26) dan como resultado las ecuaciones (1.21). Nuevamente la tercera ecuación es una ecuación desacoplada de las dos primeras.

(c) De nuevo usando (1.26) con  $\mathcal{H} = \frac{1}{2}\|p\|^2 + e^{2\alpha+\beta}f(q) + (e^\alpha - \dot{\alpha})S$  vemos que

$$\begin{aligned}\dot{q} &= p, \\ \dot{p} &= -e^{2\alpha+\beta}\nabla f(q) - p(e^\alpha - \dot{\alpha}),\end{aligned}$$

combinando estas dos ecuaciones se tiene  $\ddot{q} + (e^\alpha - \dot{\alpha})\dot{q} + e^{2\alpha+\beta}\nabla f(q) = 0$  la cual es un caso particular de la ecuación (3.31) en donde se utiliza  $h = \frac{1}{2}\|q\|^2$  y la condición (3.24). Luego se recupera la dinámica de Bregman.

(d) Usando (1.26) con  $\mathcal{H} = \frac{1}{2}\|p\|^2 + f(q) + \frac{3}{t}S$  se tiene:

$$\begin{aligned}\dot{q} &= p, \\ \dot{p} &= -\nabla f(q) - \frac{3}{t}p,\end{aligned}$$

y combinando estas dos expresiones se obtiene (3.14). ■

Debido a la proposición anterior podemos ver que también recuperamos los límites continuos obtenidos anteriormente, todo lo anterior es un caso particular de un sistema de contacto.



### 3.3. Optimización Hamiltoniana Natural

El Teorema 2.2.2 nos mostró que las funciones de divergencia inducen una métrica y dos conexiones duales, ahora veremos que ellas también inducen de manera natural una estructura simpléctica y de contacto sobre las variedades apropiadas. Para justificar la palabra **natural** recordamos que para el gradiente natural se utilizó la estructura Riemanniana “natural” de la variedad para calcular el gradiente. En esta sección utilizaremos una estructura simpléctica o de contacto “natural” de la variedad para calcular las ecuaciones de Hamilton. Tal estructura natural la construiremos usando una función de divergencia, como veremos en el siguiente teorema que se encuentra en [LZ17].

**Teorema 3.3.1.** *Sea  $M$  una variedad de dimensión  $n$  y  $D$  una función de divergencia definida sobre  $M \times M$ . Entonces  $D$  induce una forma simpléctica sobre  $M \times M$  dada por*

$$\omega_D(q, p) = - \sum_{i,j=1}^n \partial_{q_i} \partial_{p_j} D(q, p) dq_i \wedge dp_j. \quad (3.44)$$

*Demostración.* Sea  $(q, p) \in M \times M$ ,  $q$  fijo y tome la variedad  $M_p = \{q\} \times M \simeq M$  entonces  $T^*M_p$  tiene una 2-forma simpléctica que en coordenadas de Darboux  $(\eta, p)$  tiene la forma  $\omega_p = \sum_{i=1}^n dp_i \wedge d\eta_i$ . Defina la función  $R_D : M \times M \rightarrow T^*M_p$  dada por

$$R_D(q, p) = \left( p, \sum_{i=1}^n \partial_{p_i} D(q, p) dp_i \right).$$

El Jacobiano de esta función está dado por:

$$\begin{pmatrix} 0 & \partial_{p_i} \partial_{q_j} D(q, p) \\ \delta_{ij} & \partial_{p_i} \partial_{p_j} D(q, p) \end{pmatrix},$$

la cual es no degenerada en una vecindad de la diagonal  $\Delta M = \{(q, p) \in M \times M : q = p\}$ . Así  $R_D$  es localmente un difeomorfismo, luego podemos calcular el pullback de  $\omega_p$ ,

$$\begin{aligned} R_D^* \omega_p &= R_D^* \left( \sum_{i=1}^n dp_i \wedge d\eta_i \right) \\ &= \sum_{i=1}^n dp_i \wedge d(\partial_{p_i} D(q, p)) \\ &= \sum_{i=1}^n \left( dp_i \wedge \sum_{j=1}^n (\partial_{q_j} \partial_{p_i} D(q, p) dq_j + \partial_{p_j} \partial_{p_i} D(q, p) dp_j) \right) \\ &= - \sum_{i,j=1}^n \partial_{q_i} \partial_{p_j} D(q, p) dq_i \wedge dp_j, \end{aligned}$$

donde  $\sum_{j=1}^n \partial_{p_j} \partial_{p_i} D(q, p) dp_j \wedge dp_i = 0$  pues  $\partial_{p_j} \partial_{p_i} D(q, p) = \partial_{p_i} \partial_{p_j} D(q, p)$ . De manera análoga se puede definir la función  $L_D : M \times M \rightarrow T^*M_q$  dada por

$$L_D(q, p) = \left( q, \sum_{i=1}^n \partial_{q_i} D(q, p) dq_i \right),$$

de tal forma que  $L_D^* \omega_q = - \sum_{i,j=1}^n \partial_{q_i} \partial_{p_j} D(q,p) dq_i \wedge dp_j$  donde  $\omega_q = \sum_{i=1}^n dq_i \wedge d\zeta_i$  es la 2-forma simpléctica sobre  $T^*M_q$  en coordenadas de Darboux  $(q, \zeta)$ , es decir que sobre  $M \times M$  tenemos la forma simpléctica deseada. Veamos que en efecto es simpléctica, para ello veremos que la forma es cerrada y no degenerada. Para ver que es no degenerada vemos que,

$$\omega_D^n = n! \det [D_{i,j}] dq_1 \wedge dp_1 \wedge \dots \wedge dq_n \wedge dp_n,$$

la cual no se anula si  $\det [D_{i,j}] \neq 0$  y sabemos que esto no se anula en un abierto de la diagonal de  $\Delta M$ . Además, realizando un cálculo vemos que es cerrada,

$$d\omega_D = \sum_{i,j,k=1}^n \left( \frac{\partial^3 D}{\partial q_k \partial q_i \partial p_j} dq_k \wedge dq_i \wedge dp_j + \frac{\partial^3 D}{\partial p_k \partial q_i \partial p_j} dp_k \wedge dq_i \wedge dp_j \right) = 0.$$

Luego se tiene una 2 forma simpléctica. ■

Como ya tenemos una forma simpléctica el siguiente paso es calcular como quedan las ecuaciones de Hamilton utilizando la 2-forma (3.44). Note que como no estamos en coordenadas de Darboux debemos utilizar esta expresión. Para ello debemos calcular la expresión  $\iota_{X_H} \omega_D = -dH$ , luego considere las expresiones  $dH = \sum_{i=1}^n \left( \frac{\partial H}{\partial q_i} dq_i + \frac{\partial H}{\partial p_i} dp_i \right)$ ,

$X_H = \sum_{i=1}^n \left( X_{q_i} \frac{\partial}{\partial q_i} + X_{p_i} \frac{\partial}{\partial p_i} \right)$  y  $Y = \sum_{i=1}^n \left( Y_{q_i} \frac{\partial}{\partial q_i} + Y_{p_i} \frac{\partial}{\partial p_i} \right)$ . Entonces

$$\begin{aligned} \iota_{X_H} \omega_D(Y) &= \omega_D(X_H, Y) \\ &= \sum_{i,j=1}^n \left( -\partial_{p_j} \partial_{q_i} D(q,p) dq_i \wedge dp_j \right) (X_H, Y) \\ &= - \sum_{i,j=1}^n \partial_{p_j} \partial_{q_i} D(q,p) [dq_i(X_H) dp_j(Y) - dp_j(X_H) dq_i(Y)] \\ &= - \sum_{i,j=1}^n \partial_{p_j} \partial_{q_i} D(q,p) [X_{q_i} dp_j - X_{p_j} dq_i] (Y). \end{aligned}$$

Luego se tiene la relación

$$\sum_{i=1}^n \left( \sum_{j=1}^n \partial_{p_j} \partial_{q_i} D(q,p) X_{q_i} dp_j - \sum_{j=1}^n \partial_{p_j} \partial_{q_i} D(q,p) X_{p_j} dq_i \right) = \sum_{i=1}^n \left( \frac{\partial H}{\partial q_i} dq_i + \frac{\partial H}{\partial p_i} dp_i \right).$$

En el lado izquierdo al intercambiar los índices en la primera suma se tienen las expresiones

$$\frac{\partial H}{\partial q_i} = - \sum_{j=1}^n \partial_{p_j} \partial_{q_i} D(q,p) X_{p_j}, \quad (3.45)$$

$$\frac{\partial H}{\partial p_i} = \sum_{j=1}^n \partial_{q_j} \partial_{p_i} D(q,p) X_{q_j}, \quad (3.46)$$

Como  $D$  es función de divergencia sabemos que  $\partial_{p_j} \partial_{q_i} D(q,p)$  es definida positiva en un abierto de la diagonal  $\Delta M$  luego dicha matriz es invertible y se puede obtener la expresión

para  $X_H$ .

$$X_H = \sum_{i=1}^n \left( \sum_{j=1}^n [\partial_{q_j} \partial_{p_i} D(q, p)]^{-1} \frac{\partial H}{\partial p_j} \frac{\partial}{\partial q_i} - \sum_{j=1}^n [\partial_{p_j} \partial_{q_i} D(q, p)]^{-1} \frac{\partial H}{\partial q_j} \frac{\partial}{\partial p_i} \right), \quad (3.47)$$

donde  $[\partial_{p_j} \partial_{q_i} D(q, p)]^{-1}$  denota los elementos de la inversa de la matriz  $(\partial_{p_j} \partial_{q_i} D(q, p))_{i,j=1}^n$ . Así las ecuaciones de Hamilton son:

$$\begin{cases} \dot{q}_i = \sum_{j=1}^n [\partial_{q_j} \partial_{p_i} D(q, p)]^{-1} \frac{\partial H}{\partial p_j} \\ \dot{p}_i = - \sum_{j=1}^n [\partial_{p_j} \partial_{q_i} D(q, p)]^{-1} \frac{\partial H}{\partial q_j}. \end{cases} \quad (3.48)$$

**Ejemplo 3.3.2.** Como primer ejemplo considere  $D(q, p) = A_F(q, p)$  la divergencia canónica entonces por el inciso (c) del Ejemplo 2.2.5 sabemos que  $-\partial_{q_i} \partial_{p_j} D(q, p) = \delta_{i,j}$ , por tanto se obtiene  $D^{j,i}(q, p) = \delta^{j,i}$  y  $D^{i,j}(q, p) = \delta^{i,j}$ , así el campo Hamiltoniano tiene la forma:

$$X_H = \sum_{i=1}^n \left( \frac{\partial H}{\partial p_i} \frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_i} \frac{\partial}{\partial p_i} \right).$$

Luego en este caso se obtiene el campo Hamiltoniano (1.11) que ya habíamos obtenido. Si además  $H(q, p) = A_F(q, p)$  y usando las ecuaciones (2.20) y (2.21) se tiene

$$\begin{cases} \dot{q}_i = \partial_{p_i} F^*(p) - q_i \\ \dot{p}_i = -\partial_{q_i} F(q) + p_i. \end{cases} \quad (3.49)$$

**Ejemplo 3.3.3.** También podemos usar  $D(q, p) = B_F(q, p)$  la divergencia de Bregman, donde por el inciso (c) del Ejemplo 2.2.3 se tiene  $-\partial_{q_j} \partial_{p_i} B_F(q, p) = \partial_{p_j} \partial_{p_i} F(p) = (\nabla^2 F(p))_{i,j}$ . Denotamos  $(\nabla^2 F(p))_{i,j}^{-1}$  la inversa de la matriz Hessiana (la cual existe porque la matriz es definida positiva y por tanto invertible), entonces,

$$\begin{aligned} X_H &= \sum_{i=1}^n \left( \sum_{j=1}^n (\nabla^2 F(p))_{j,i}^{-1} \frac{\partial H}{\partial p_j} \frac{\partial}{\partial q_i} - \sum_{j=1}^n (\nabla^2 F(p))_{i,j}^{-1} \frac{\partial H}{\partial q_j} \frac{\partial}{\partial p_i} \right) \\ &= \sum_{i,j=1}^n (\nabla^2 F(p))_{i,j}^{-1} \left( \frac{\partial H}{\partial p_j} \frac{\partial}{\partial q_i} - \frac{\partial H}{\partial q_j} \frac{\partial}{\partial p_i} \right), \end{aligned}$$

de donde se obtiene que las ecuaciones de Hamilton para este caso toman la forma:

$$\begin{cases} \dot{q}_i = \sum_{j=1}^n (\nabla^2 F(p))_{i,j}^{-1} \frac{\partial H}{\partial p_j}, \\ \dot{p}_i = - \sum_{j=1}^n (\nabla^2 F(p))_{i,j}^{-1} \frac{\partial H}{\partial q_j}. \end{cases} \quad (3.50)$$

Si además  $H(q, p) = B_F(q, p)$  y usando las ecuaciones (2.12) y (2.13) se tiene

$$\begin{cases} \dot{q}_i = - \sum_{j=1}^n (\nabla^2 F(p))_{i,j}^{-1} \langle q - p, \partial_{p_i} \nabla F(p) \rangle, \\ \dot{p}_i = - \sum_{j=1}^n (\nabla^2 F(p))_{i,j}^{-1} (\partial_{q_i} F(q) - \partial_{p_i} F(p)). \end{cases} \quad (3.51)$$

Note que en ambos ejemplos se obtiene una 2-forma simpléctica (la cuál es no degenerada) en todo  $M \times M$  y no solo cerca de la diagonal.

El Teorema 3.3.1 nos da una estructura simpléctica “natural”, sin embargo, esta no es útil para hacer optimización, pues las ecuaciones de Hamilton obtenidas son conservativas. Por otro lado, vimos que la estructura de contacto incluye los casos de Hamiltonianos dependientes del tiempo y sistemas conformalmente simplécticos, luego es natural preguntarnos si el Teorema 3.3.1 tiene un análogo para el caso de una estructura de contacto y poder trabajar con ella en optimización. La respuesta a esto nos lo da el siguiente teorema:

**Teorema 3.3.4.** *Sea  $M$  una variedad de dimensión  $n$  y  $D$  una función de divergencia definida sobre  $M \times M$ . Entonces  $D$  induce dos formas de contacto sobre  $M \times M \times \mathbb{R}$  dadas por*

$$\alpha_D^1(q, p, S) = dS - \sum_{i=1}^n \partial_{q_i} D(q, p) dq_i, \quad (3.52)$$

$$\alpha_D^2(q, p, S) = dS - \sum_{i=1}^n \partial_{p_i} D(q, p) dp_i. \quad (3.53)$$

*Demostración.* Al igual que en la prueba del Teorema 3.3.1 consideramos la notación  $M_p = \{q\} \times M \simeq M$  y  $M_q = M \times \{p\} \simeq M$ . Defina la función  $L_D : M \times M \times \mathbb{R} \rightarrow T^*M_q \times \mathbb{R}$  dada por

$$L_D(q, p, S) = (q, \sum \partial_{q_i} D(q, p) dq_i, S).$$

El Jacobiano de esta función está dado por

$$\begin{pmatrix} \delta_{i,j} & \partial_{q_i} \partial_{q_j} D(q, p) & 0 \\ 0 & \partial_{p_j} \partial_{q_i} D(q, p) & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

la cual es no degenerada en una vecindad de la diagonal  $\Delta M$ . Así  $L_D$  es localmente un difeomorfismo. Sobre  $T^*M_q \times \mathbb{R}$  en coordenadas de Darboux  $(q, \xi, S)$  podemos escribir la forma de contacto como  $\alpha_q = dS - \sum \xi_i dq_i$ , así calculando el pullback se tiene,

$$\begin{aligned} L_D^* \alpha_q &= L_D^* \left( dS - \sum \xi_i dq_i \right) \\ &= L_D^* (dS) - \sum L_D^* (\xi_i dq_i) \\ &= dS - \sum \partial_{q_i} D(q, p) dq_i. \end{aligned}$$

De manera análoga se puede definir la función  $R_D : M \times M \times \mathbb{R} \rightarrow T^*M_p \times \mathbb{R}$  definida como

$$R_D(q, p, S) = (p, \sum \partial_{p_i} D(q, p) dp_i, S).$$

De tal forma que sobre  $T^*M_p \times \mathbb{R}$  tenemos una forma de contacto, que en coordenadas de Darboux  $(p, \eta, S)$  esta dada por  $\alpha_p = dS - \sum \eta_i dp_i$  y así calculando el pullback se obtiene  $R_D^* \alpha_p = dS - \sum \partial_{p_i} D(q, p) dp_i$ . Por otro lado, note que que la derivada exterior de ambas 1-formas dan como resultado la ecuación (3.44) (módulo un signo), por tanto efectivamente son formas de contacto, es más, ambas 1-formas obtenidas son la contactificación de las

2 1-formas exactas que se hubieran obtenido si en el Teorema 3.3.1 hubiéramos hecho el pullback de la 1-forma tautológica en lugar de hacerlo con la 2-forma. Así sobre  $M \times M \times \mathbb{R}$  tenemos dos formas de contacto. ■

Ahora vamos a calcular las ecuaciones de Hamilton para las 1-formas de contacto (3.52) y (3.53). Note que como no estamos en coordenadas de Darboux debemos utilizar estas expresiones. Para ello tenemos que usar las expresiones de la Definición 1.3.4. Primero, note que el campo de Reeb para ambas 1-formas es  $\xi = \frac{\partial}{\partial S}$  pues un cálculo permite comprobar que  $\iota_\xi d\alpha_D^j = 0$  y  $\iota_\xi \alpha_D^j = 1$  para  $j \in \{1, 2\}$ . Así por el Lema 1.3.5 se tiene  $f_{\mathcal{H}} = -\frac{\partial \mathcal{H}}{\partial S}$ . Considere  $X_{\mathcal{H}} = \sum_{i=1}^n \left( X_{q_i} \frac{\partial}{\partial q_i} + X_{p_i} \frac{\partial}{\partial p_i} \right) + X_S \frac{\partial}{\partial S}$  y además tenemos que  $d\mathcal{H} = \sum_{i=1}^n \left( \frac{\partial \mathcal{H}}{\partial q_i} dq_i + \frac{\partial \mathcal{H}}{\partial p_i} dp_i \right) + \frac{\partial \mathcal{H}}{\partial S} dS$ . Primero usamos la relación  $-\mathcal{H} = \iota_{X_{\mathcal{H}}} \alpha_D^1$  entonces

$$-\mathcal{H} = \iota_{X_{\mathcal{H}}} \alpha_D^1 = \alpha_D^1(X_{\mathcal{H}}) = X_S - \sum_{i=1}^n \partial_{q_i} D(q, p) X_{q_i}.$$

Luego se tiene:

$$X_S = \sum_{i=1}^n \partial_{q_i} D(q, p) X_{q_i} - \mathcal{H}. \quad (3.54)$$

Por otro lado, usando la identidad de Cartan se tiene

$$\begin{aligned} \mathcal{L}_{X_{\mathcal{H}}} \alpha_D^1 &= \iota_{X_{\mathcal{H}}} (d\alpha_D^1) - d\mathcal{H} \\ &= \iota_{X_{\mathcal{H}}} \left( \sum_{i=1}^n dq_i \wedge d(\partial_{q_i} D(q, p)) \right) - d\mathcal{H} \\ &= \iota_{X_{\mathcal{H}}} \left( \sum_{i,j=1}^n \partial_{p_j} \partial_{q_i} D(q, p) dq_i \wedge dp_j \right) - d\mathcal{H} \\ &= \sum_{i=1}^n \left( \sum_{j=1}^n \partial_{q_j} \partial_{p_i} D(q, p) X_{q_j} dp_i - \sum_{j=1}^n \partial_{p_j} \partial_{q_i} D(q, p) X_{p_j} dq_i \right) - d\mathcal{H}. \end{aligned}$$

Esta última expresión es igual a

$$= \sum_{i=1}^n \left( \sum_{j=1}^n \left( \partial_{q_j} \partial_{p_i} D(q, p) X_{q_j} - \frac{\partial \mathcal{H}}{\partial p_i} \right) dp_i - \sum_{j=1}^n \left( \partial_{p_j} \partial_{q_i} D(q, p) X_{p_j} + \frac{\partial \mathcal{H}}{\partial q_i} \right) dq_i \right) - \frac{\partial \mathcal{H}}{\partial S} dS.$$

Pero esto debe ser igual a  $f_{\mathcal{H}} \alpha_D^1 = -\frac{\partial \mathcal{H}}{\partial S} \alpha_D^1 = -\frac{\partial \mathcal{H}}{\partial S} dS + \sum_{i=1}^n \frac{\partial \mathcal{H}}{\partial S} \partial_{q_i} D(q, p) dq_i$ . De acá se obtienen las expresiones:

$$\sum_{j=1}^n \partial_{p_i} \partial_{q_j} D(q, p) X_{q_j} = \frac{\partial \mathcal{H}}{\partial p_i},$$

$$\sum_{j=1}^n \partial_{p_j} \partial_{q_i} D(q, p) X_{p_j} + \frac{\partial \mathcal{H}}{\partial q_i} = -\frac{\partial \mathcal{H}}{\partial S} \partial_{q_i} D(q, p).$$

Luego las componentes del campo Hamiltoniano son:

$$X_{q_i} = \sum_{j=1}^n [\partial_{p_i} \partial_{q_j} D(q, p)]^{-1} \frac{\partial \mathcal{H}}{\partial p_j}, \quad (3.55)$$

$$X_{p_i} = - \sum_{j=1}^n [\partial_{p_j} \partial_{q_i} D(q, p)]^{-1} \left( \frac{\partial \mathcal{H}}{\partial S} \partial_{q_j} D(q, p) + \frac{\partial \mathcal{H}}{\partial q_j} \right), \quad (3.56)$$

$$X_S = \sum_{i,j=1}^n \partial_{q_i} D(q, p) [\partial_{p_i} \partial_{q_j} D(q, p)]^{-1} \frac{\partial \mathcal{H}}{\partial p_j} - \mathcal{H}, \quad (3.57)$$

donde (3.57) se obtiene de reemplazar (3.55) en (3.54). Un cálculo similar usando  $\alpha_D^2$  nos da que en este caso las componentes del campo Hamiltoniano tienen la forma:

$$X_{q_i} = \sum_{j=1}^n [\partial_{p_j} \partial_{q_i} D(q, p)]^{-1} \left( \frac{\partial \mathcal{H}}{\partial S} \partial_{p_j} D(q, p) + \frac{\partial \mathcal{H}}{\partial p_j} \right), \quad (3.58)$$

$$X_{p_i} = - \sum_{j=1}^n [\partial_{p_i} \partial_{q_j} D(q, p)]^{-1} \frac{\partial \mathcal{H}}{\partial q_j}, \quad (3.59)$$

$$X_S = - \sum_{i,j=1}^n \partial_{p_i} D(q, p) [\partial_{p_i} \partial_{q_j} D(q, p)]^{-1} \frac{\partial \mathcal{H}}{\partial q_j} - \mathcal{H}. \quad (3.60)$$

**Ejemplo 3.3.5.** Como primer ejemplo consideraremos  $D(q, p) = A_F(q, p)$  la divergencia canónica y usando (3.52), se tiene

$$\begin{cases} \dot{q}_i = - \frac{\partial \mathcal{H}}{\partial p_i} \\ \dot{p}_i = \frac{\partial \mathcal{H}}{\partial S} (\partial_{q_i} F(q) - p_i) + \frac{\partial \mathcal{H}}{\partial q_i} \\ \dot{S} = - \sum_i^n (\partial_{q_i} F(q) - p_i) \frac{\partial \mathcal{H}}{\partial p_i} - \mathcal{H}, \end{cases} \quad (3.61)$$

las cuales si comparamos con las ecuaciones (1.26) se parecen salvo el factor extra que está apareciendo  $\partial_{q_i} F(q)$ . En el caso que  $\mathcal{H}(q, p, S) = A_F(q, p) + \gamma S$  con  $\gamma \in \mathbb{R}$  se tiene:

$$\begin{cases} \dot{q}_i = q_i - \partial_{p_i} F^*(p) \\ \dot{p}_i = (\gamma + 1) (\partial_{q_i} F(q) - p_i) \\ \dot{S} = \sum_i^n (\partial_{q_i} F(q) - p_i) (\partial_{p_i} F^*(p) - q_i) - F(q) - F^*(p) + \langle q, p \rangle - \gamma S. \end{cases} \quad (3.62)$$

**Ejemplo 3.3.6.** Ahora si  $D(q, p) = B_F(q, p)$  las ecuaciones de Hamilton de contacto toman la forma

$$\begin{cases} \dot{q}_i = \sum_{j=1}^n (\nabla^2 F(p))_{i,j}^{-1} \frac{\partial \mathcal{H}}{\partial p_j} \\ \dot{p}_i = \sum_{j=1}^n (\nabla^2 F(p))_{i,j}^{-1} \left( \frac{\partial \mathcal{H}}{\partial S} (\partial_{q_j} F(q) - \partial_{p_j} F(p)) - \frac{\partial \mathcal{H}}{\partial q_j} \right) \\ \dot{S} = \sum_{i,j=1}^n (\partial_{q_j} F(q) - \partial_{p_j} F(p)) (\nabla^2 F(p))_{i,j}^{-1} \frac{\partial \mathcal{H}}{\partial p_j} - \mathcal{H}. \end{cases} \quad (3.63)$$

En el caso que  $\mathcal{H}(q, p, S) = B_F(q, p) + \gamma S$  con  $\gamma \in \mathbb{R}$  se tiene

$$\begin{cases} \dot{q}_i = - \sum_{j=1}^n (\nabla^2 F(p))_{i,j}^{-1} \langle q - p, \partial_{p_j} \nabla F(p) \rangle \\ \dot{p}_i = (\gamma - 1) \sum_{j=1}^n (\nabla^2 F(p))_{i,j}^{-1} (\partial_{q_j} F(q) - \partial_{p_j} F(p)) \\ \dot{S} = - \sum_{i,j=1}^n (\partial_{q_j} F(q) - \partial_{p_j} F(p)) (\nabla^2 F(p))_{i,j}^{-1} \langle q - p, \partial_{p_j} \nabla F(p) \rangle - B_F(q, p) - \gamma S. \end{cases}$$

Usando lo anterior, queremos buscar un resultado análogo al teorema principal de [FJV20] que enunciamos en apéndice (Teorema A.0.11), luego necesitamos lo siguiente:

**Lema 3.3.7.** *Sea  $H$  un Hamiltoniano cuyas ecuaciones del movimiento están dadas por (1.12). Considere una modificación del flujo Hamiltoniano (1.12) dada por*

$$\begin{cases} \dot{q} = \nabla_p H + q^* - q \\ \dot{p} = -\nabla_q H + p^* - p, \end{cases} \quad (3.64)$$

donde  $(q^*, p^*)$  es el punto donde  $H$  alcanza su mínimo. Entonces si  $H$  es una función convexa en  $q$  y en  $p$ , tal que  $H(q^*, p^*) = 0$  y  $H \geq 0$  se cumple la desigualdad

$$H(q(t), p(t)) \leq H(q(0), p(0))e^{-t}. \quad (3.65)$$

*Demostración.* Note que

$$\begin{aligned} \dot{H}(q, p) &= \langle \nabla_q H, \dot{q} \rangle + \langle \nabla_p H, \dot{p} \rangle \\ &= \langle \nabla_q H, \nabla_p H + q^* - q \rangle + \langle \nabla_p H, -\nabla_q H + p^* - p \rangle \\ &= \langle \nabla_q H, q^* - q \rangle + \langle \nabla_p H, p^* - p \rangle \\ &= \begin{pmatrix} \nabla_q H & \nabla_p H \end{pmatrix} \begin{pmatrix} q^* - q \\ p^* - p \end{pmatrix} \\ &\leq H(q^*, p^*) - H(q, p) = -H(q, p), \end{aligned}$$

donde la desigualdad se tiene porque  $H$  es convexa. Luego  $\dot{H}(q, p) \leq -H(q, p)$  y de acá por la desigualdad de Grönwall se tiene el resultado. ■

Ahora considere el Hamiltoniano

$$H(q, p) = B_f(q, q^*) + B_{f^*}(p, p^*). \quad (3.66)$$

Note que  $p^* = \nabla_q f(q^*) = 0$  y por tanto se tiene,

$$B_f(q, q^*) = f(q) - f(q^*). \quad (3.67)$$

Además, note que  $H(q^*, p^*) = 0$ . Si utilizamos este Hamiltoniano en la ecuación (3.64) y usando que  $\nabla_p f^*(p^*) = q^*$  se obtiene

$$\begin{cases} \dot{q} = \nabla_p f^*(p) - q \\ \dot{p} = -\nabla_q f(q) - p, \end{cases} \quad (3.68)$$

El problema con el Hamiltoniano (3.66) es que depende del punto óptimo que queremos encontrar, así que vamos a definir un Hamiltoniano de contacto que no dependa de  $(q^*, p^*)$ . Considere el Hamiltoniano  $H$  dado por (3.66) y construyamos  $\mathcal{H}$  de la siguiente forma:

$$\mathcal{H} = -H + \langle q, p \rangle - \langle q^*, p \rangle + \langle q, p^* \rangle + 2S - 2f(q) \quad (3.69)$$

$$= -f(q) - f^*(p) + \langle q, p \rangle + 2S - 2f(q) \quad (3.70)$$

$$= -A_f(q, p) + 2S - 2f(q), \quad (3.71)$$

donde vemos que la última igualdad no depende del punto óptimo.

Para encontrar el sistema Hamiltoniano usaremos las ecuaciones de Hamilton inducidas por la divergencia canónica dada por la ecuación (3.61), al realizar el cálculo vemos que el sistema que se obtiene es el (3.68) y por lo tanto obtenemos el siguiente resultado.

**Teorema 3.3.8.** *El sistema Hamiltoniano de contacto dado por el Hamiltoniano (3.71) junto con las ecuaciones de contacto (3.61) converge exponencialmente al punto óptimo, es decir,*

$$f(q(t)) - f(q^*) \leq e^{-t}(f(q(0)) - f(q^*)). \quad (3.72)$$

*Demostración.* Como (3.66) y (3.71) generan el mismo sistema Hamiltoniano (3.68) entonces si probamos la desigualdad (3.72) usando (3.66) entonces también funciona para (3.71). Como (3.66) satisface el Lema 3.3.7 entonces se tiene

$$B_f(q(t), q^*) + B_{f^*}(p(t), p^*) = H(q(t), p(t)) \leq H(q(0), p(0))e^{-t}$$

Colocando  $p(0) = p^* = 0$  y usando que  $B_{f^*}(p(t), p^*) \geq 0$  y (3.67) se tiene

$$f(q(t)) - f(q^*) \leq e^{-t}(B_f(q(0), q^*) + B_{f^*}(p(0), p^*)) = e^{-t}(f(q(0)) - f(q^*))$$

Luego se tiene la desigualdad deseada. ■

A continuación, vamos a agregar algunos comentarios sobre el Teorema 3.3.8, que es el resultado principal del presente trabajo.

- Note que en la demostración no se utilizó ninguna propiedad específica de  $f^*$ , por lo tanto en general este teorema funciona con Hamiltonianos de la forma

$$\mathcal{H} = -B_f(q, q^*) - B_{F^*}(p, p^*) + \langle q, p \rangle - \langle q^*, p \rangle + \langle q, p^* \rangle + 2S - 2f(q), \quad (3.73)$$

donde  $F$  es cualquier otra función suave y convexa que debe cumplir que  $q^* = \nabla F^*(p)$ .

- Hasta el momento en la literatura el único resultado para una convergencia exponencial hacia el óptimo está dado por el Hamiltoniano de Bregman definido en la Sección 3.2.1, Ecuación (3.22). Por un lado, hay un parecido entre el Hamiltoniano de Bregman y el encontrado en la ecuación (3.71). Por otro lado, nuestra derivación es más general y más geométrica, y la forma del Hamiltoniano en (3.71) permite hacer una descomposición directa de su flujo para obtener una discretización que respete las propiedades del flujo continuo (ver el apéndice).



- Note que para llegar al Teorema [3.3.8](#) se utilizó tanto la divergencia canónica (para inducir la 1-forma de contacto apropiada sobre  $M \times M \times \mathbb{R}$ ) como la divergencia de Bregman (para encontrar al Hamiltoniano). Entonces vemos como las estructuras que definen la geometría de la información juegan un papel importante en este caso.
- A pesar de haber utilizado a la divergencia canónica para inducir a la 1-forma de contacto, no hemos utilizado la estructura de variedad dualmente plana que está también inducida por la misma divergencia (ver Ejemplo [2.2.5](#)). Es posible que se pueda encontrar otros algoritmos de optimización que utilicen las geodésicas de las conexiones duales planas inducidas por la divergencia canónica.
- Con respecto a la expresión [\(3.73\)](#), note que

$$\begin{aligned}
 B_{F^*}(p, p^*) &= F^*(p) - F^*(p^*) - \langle p - p^*, \nabla F^*(p^*) \rangle \\
 &= F^*(p) - F^*(p^*) - \langle p - p^*, q^* \rangle \\
 &= F^*(p) - F^*(p^*) - \langle p, q^* \rangle
 \end{aligned}$$

entonces el término  $\langle p, q^* \rangle$  siempre desaparece en [\(3.73\)](#) y recuerde que  $p^* = 0$ , así que el Hamiltoniano [\(3.73\)](#) tampoco depende de la solución óptima  $q^*, p^*$ .

En este trabajo se exploró la relación entre la geometría de la mecánica y la geometría de la información. Para ello hemos revisado algunos de los conceptos generales de la geometría simpléctica y de contacto, con los cuales se pueden definir geoméricamente sistemas mecánicos con disipación, y después hemos revisado algunas de las herramientas típicas de la geometría de la información, como por ejemplo las funciones de divergencia que inducen estructuras métricas y conexiones duales sobre una variedad. Finalmente, se ha probado que una función de divergencia induce también una estructura simpléctica o de contacto “natural” sobre la variedad apropiada (Teoremas [3.3.1](#) y [3.3.4](#)), así estableciendo un enlace concreto entre la geometría subyacente a la mecánica y la geometría de la información.

Por otro lado, hemos utilizado el estudio mencionado anteriormente para encontrar nuevos sistemas dinámicos Hamiltonianos cuya evolución resuelva el problema de encontrar el mínimo de una función convexa asignada. Este resultado (Teorema [3.3.8](#)) es el resultado principal del presente trabajo. El primer aspecto interesante del resultado obtenido es que a nivel continuo la tasa de convergencia obtenida es exponencial, lo cual extiende y simplifica algunos resultados anteriores en la literatura [\[WWJ16\]](#). Otro aspecto relevante del Teorema [3.3.8](#) es que para obtenerlo se han usado tanto la divergencia canónica, la cual nos ha permitido encontrar la 1-forma de contacto “natural” con la cual trabajar, como la divergencia de Bregman, la cual nos ha ayudado a definir el Hamiltoniano, probando así que el enlace anteriormente establecido entre geometría de la mecánica y geometría de la información puede ser muy fructuoso.

Este trabajo deja abiertas diferentes direcciones de investigación. Para empezar, no hemos profundizado el estudio de las estructuras simplécticas y de contacto inducidas por las funciones de divergencia y de su relación con las estructuras métricas y de conexiones duales que también son inducidas por las mismas divergencias. Algunos resultados en este sentido pueden encontrarse en [\[ZL16\]](#) y [\[ZL13\]](#).

Para la parte de optimización, queda abierta una dirección fundamental, que es proponer una discretización del flujo Hamiltoniano de contacto correspondiente al Teorema [3.3.8](#) y encontrar el orden de convergencia del correspondiente algoritmo discreto. Para ello, se espera que exista un resultado análogo al Teorema [A.0.11](#) obtenido en [\[FJV20\]](#), donde se prueba que los algoritmos de discretización geométricos brevemente resumidos en el

apéndice respetan el orden de convergencia continuo salvo por un error que decrece exponencialmente en el tiempo.

Otras direcciones en las que se pueden extender los resultados de este trabajo son:

1. Utilizar la estructura geométrica simpléctica o de contacto para probar más propiedades de los sistemas Hamiltonianos de interés en optimización, por ejemplo, ¿se puede encontrar una equivalencia módulo una transformación de contacto dependiente del tiempo del Hamiltoniano de Bregman (3.22) con el Hamiltoniano del Teorema 3.3.8?, ¿se puede usar a los contactomorfismos y a las simetrías de los sistemas que se pretende optimizar para reducir el problema de optimización? (como en [LL19]), ¿se puede considerar problemas de optimización con restricciones en un esquema geométrico de tipo Dirac? (como en [LGLRR20]).
2. En principio se pueden considerar otras funciones de divergencia y otros tipos de Hamiltonianos para encontrar diferentes algoritmos de optimización, y quizás poder abordar el caso de funciones no convexas.
3. Estaría interesante extender el análisis desarrollado en [LZ17] sobre la equivalencia entre las funciones de divergencia y los llamados “Lagrangianos discretos exactos” en la teoría geométrica de discretización de sistemas Hamiltonianos al caso de contacto. Esto podría llevar a otro tipo de enfoque más “variacional” con respecto al problema de optimización.
4. Se podría tratar de extender parte del análisis de este trabajo a otros temas de recién interés en Machine Learning como por ejemplo los llamados “Hamiltonian Neural Networks” (ver [GDY19]).

- [A16] Amari, S. *Information Geometry and its Applications*. Springer, 2016.
- [BJW18] Betancourt M., Jordan M. I. & Wilson A. C. *On Symplectic Optimization*. arXiv:1802.03653 [stat.CO], 2018.
- [BCT17] Bravetti A., Cruz H. & Tapias D. *Contact Hamiltonian Mechanics*. Annals of Physics, 376: 17 – 39, 2017.
- [BDFB19] Bravetti A., Daza M. L., Flores H. & Betancourt M. *Optimization algorithms inspired by the geometry of dissipative systems*. arXiv:1912.029028 [math.OC], 2019.
- [BSVZ19] Bravetti A., Seri M., Vermeeren M. & Zadra F. *Numerical Integration in Celestial Mechanics: A Case for Contact Geometry*. Celest Mech Dyn Astr 132, 7 (2020).
- [dS01] da Silva A.C. *Lectures on Symplectic Geometry*. Springer, 2001.
- [FJV20] França G., Jordan M. I. & Vidal R. *On Dissipative Symplectic Integration With Applications to Gradient-Based Optimization*. arXiv:2004.06840 [math.OC], 2020.
- [FSRV19] França G., Sulam J., Robinson D. P., Vidal R., *Conformal Symplectic and Relativistic Optimization*. arXiv:1903.04100 [math.OC], 2019.
- [G01] Geiges H. *A brief history of contact geometry and topology*. Expositiones Mathematicae. Volume 19, Issue 1, 2001, Pages 25-53.
- [GDY19] Greydanus S., Dzamba M., Yosinski J. *Hamiltonian neural networks*. CoRR, abs/1906.01563, 2019.
- [HLW06] Hairer E., Lubich C., & Wanner G. *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*. Volume 31 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin Heidelberg, 2006.
- [LZ17] Leok M. & Zhang J. *Connecting information geometry and geometric mechanics*. Entropy, 19:518, 2017.
- [LGLRR20] León M., Gaset J., Laínz M., Rivas X., Román-Roy N. *Unified Lagrangian-Hamiltonian formalism for contact systems*. arXiv:2003.13037 [math-ph], 2020.

- [LL19] Leon M., Lainz Valcazar M. *Contact Hailtonian Systems* Journal of Mathematical Physics 60, 102902 (2019).
- [L19] Liu Y. *An Overview of Information Geometry*. 2019. Recuperado de: <https://yuxiliu1995.github.io/notes/>
- [Mc15] McInerney A. *First Steps in Differential Geometry*. Springer, 2015.
- [McP01] McLachlan R. & Perlmutter M. *Conformal Hamiltonian systems*. Journal of Geometry and Physics 39 (2001) 276-300
- [N18] Nielsen F. *An elementary introduction to information geometry*. CoRR, abs/1808.08271. 2018.
- [O'DM19] O'Donoghue B. & Maddison C. J. *Hamiltonian descent for composite objectives*. In Advances in Neural Information Processing Systems, pages 14443–14453, 2019.
- [SBC16] Su W., Boyd S. & Candès E. J. *A Differential Equation for Modeling Nesterov's Accelerated Gradient Method: Theory and Insights*. arXiv:1503.01243, [stat.ML], 2015.
- [SMDH13] Sutskever I., Martens J., Dahl G. E. & Hinton G. E. *On the importance of initialization and momentum in deep learning*. Proceedings of the 30th International Conference on International Conference on Machine Learning - Volume 28 June 2013.
- [WWJ16] Wibisono A., Wilson A. C. & Jordan M. I. *A variational perspective on accelerated methods in optimization*. Proceedings of National Academy of Science, 113(47) : E7351 – E7358, 2016.
- [Y90] Yoshida H. *Construction of higher order symplectic integrators*. Physics letters A, 150:262-268, 1990.
- [ZL16] Zhang J, Fei T. *Information Geometry with (Para-)Kähler Structures*. En: Ay N., Gibilisco P., Matúš F. (eds) *Information Geometry and Its Applications*. IGAIA IV 2016. Springer Proceedings in Mathematics & Statistics, vol 252. Springer, Cham.
- [ZL13] Zhang J, Li F. *Symplectic and Kähler Structures on Statistical Manifolds Induced from Divergence Functions*. En: Nielsen F., Barbaresco F. (eds) *Geometric Science of Information*. GSI 2013. Lecture Notes in Computer Science, vol 8085. Springer, Berlin, Heidelberg.

La integración geométrica es la integración numérica de una ecuación diferencial que preserva una o más de sus propiedades geométricas. Ejemplo de las propiedades geométricas son volumen, energía, estructura simpléctica, entre otros.

Comenzamos definiendo que es un integrador numérico, veremos los que son de nuestro interés, los simplécticos y así introducimos ejemplos del método de Euler simpléctico y no simpléctico. Además, vemos que cuando un campo vectorial no es integrable lo podemos descomponer en campos que si sean integrables y así poder construir de manera más fácil integradores simplécticos o de contacto. Damos ejemplos de esto y concluimos con un teorema de convergencia que justifica porque trabajamos con sistemas continuos y no discretos. Para exponer esto nos guiaremos de [HLW06], [BSVZ19] y [EJV20].

**Definición A.0.1.** Un *integrador numérico* es una función  $\widehat{\Phi}_\tau : \mathbb{R}^{2n} \rightarrow \mathbb{R}^{2n}$  con paso  $\tau > 0$  tal que las iteraciones

$$x_k = \widehat{\Phi}_\tau(x_{k-1}), \quad x(0) = x_0, \quad (\text{A.1})$$

aproximan el verdadero estado del sistema,  $x(t_k) = (q(t_k), p(t_k))$  en un instante discreto  $t_k = \tau k$  para  $k \in \mathbb{N}$ . Además, si  $\Phi_t$  es el flujo verdadero entonces el integrador se dice de **orden**  $r \geq 1$  si

$$\left\| \widehat{\Phi}_\tau(x) - \Phi_t(x) \right\| = \mathcal{O}(\tau^{r+1}).$$

**Definición A.0.2.** Un *integrador numérico*  $\widehat{\Phi}_\tau$ , donde  $\tau$  es el tamaño de paso, para un sistema Hamiltoniano conservativo se llama un **integrador simpléctico** si preserva la estructura simpléctica, es decir,  $\widehat{\Phi}_\tau^* \omega = \omega$ . Además, se dice **simpléctico conforme** si  $\widehat{\Phi}_\tau$  es conforme simpléctico, es decir,  $\widehat{\Phi}_\tau^* \omega_{k+1} = e^{-c\tau} \omega_k$  siempre y cuando  $\widehat{\Phi}_h$  sea aplicado a un campo Hamiltoniano suave. Note que iterando  $k$  veces se obtiene  $\widehat{\Phi}_\tau^* \omega_k = e^{-c\tau k} \omega_0$ .

**Ejemplo A.0.3.** Consideremos el método de Euler para el Hamiltoniano del oscilador armónico adimensional  $H = \frac{q^2 + p^2}{2}$ . Las ecuaciones de Hamilton toman la forma  $\frac{dq}{dt} = p$  y  $\frac{dp}{dt} = -q$  aplicando el método de Euler (no simpléctico) se tiene

$$q_{k+1} = q_k + \tau p_k \quad (\text{A.2})$$

$$p_{k+1} = p_k - \tau q_k, \quad (\text{A.3})$$

respectivamente para cada ecuación. Por otro lado, note que

$$\begin{aligned}
 H_{k+1} &= \frac{q_{k+1}^2 + p_{k+1}^2}{2} \\
 &= \frac{1}{2} \left( (q_k + \tau p_k)^2 + (p_k - \tau q_k)^2 \right) \\
 &= \frac{1}{2} \left( q_k^2 + p_k^2 + \tau^2 (p_k^2 + q_k^2) \right) \\
 &= H_k + \tau^2 H_k.
 \end{aligned}$$

Esto nos permite ver que el error de la energía se va acumulando en cada paso  $\tau$  pues el término  $\tau^2 H_k$  siempre es positivo.

**Ejemplo A.0.4.** Ahora consideraremos el método simpléctico de Euler para el Hamiltoniano del oscilador armónico adimensional  $H = \frac{q^2 + p^2}{2}$ . Las ecuaciones de Hamilton toman la forma  $\frac{dq}{dt} = p$  y  $\frac{dp}{dt} = -q$  aplicando el método simpléctico de Euler se tiene

$$q_{k+1} = q_k + \tau p_k \quad (\text{A.4})$$

$$p_{k+1} = p_k - \tau q_{k+1}, \quad (\text{A.5})$$

respectivamente para cada ecuación. Por otro lado, note que

$$\begin{aligned}
 H_{k+1} &= \frac{q_{k+1}^2 + p_{k+1}^2}{2} \\
 &= \frac{1}{2} \left( (q_k + \tau p_k)^2 + (p_k - \tau q_{k+1})^2 \right) \\
 &= \frac{1}{2} \left( q_k^2 + 2\tau q_k p_k + \tau^2 p_k^2 + p_k^2 - 2\tau p_k (q_k + \tau p_k) + \tau^2 (q_k + \tau p_k)^2 \right) \\
 &= \frac{1}{2} \left( q_k^2 + \tau^2 p_k^2 + p_k^2 - 2\tau^2 p_k^2 + \tau^2 (q_k^2 + 2\tau q_k p_k + \tau^2 p_k^2) \right) \\
 &= \frac{1}{2} \left( q_k^2 + p_k^2 + \tau^2 (q_k^2 + p_k^2) - 2\tau^2 p_k^2 + 2\tau^3 q_k p_k + \tau^4 p_k^2 \right) \\
 &= H_k + \tau^2 H_k - \tau^2 p_k^2 + \tau^3 q_k p_k + \frac{1}{2} \tau^4 p_k^2
 \end{aligned}$$

Sin pérdida de generalidad podemos omitir los términos de  $\tau^3$  y  $\tau^4$  así se tiene la relación  $H_{k+1} = H_k + \tau^2 (H_k - p_k^2)$  donde podemos ver que a diferencia del método de Euler el error no se acumula, si no que va oscilando alrededor del valor del Hamiltoniano, pues el término  $H_k - p_k^2$  puede ser positivo o negativo.

**Definición A.0.5.** Sea  $X$  campo vectorial. Decimos que  $X$  es exactamente integrable si la ecuación diferencial  $\dot{x} = X(x)$ , con problema de valor inicial  $x(0) = x_0$ , tiene solución  $x(t) = \exp(tX)x_0$  donde  $\exp(tX)$  se puede encontrar analíticamente como el operador flujo del campo vectorial en un tiempo  $t$ .

**Ejemplo A.0.6.** Considere la ecuación  $\left(\frac{dx}{dt}, \frac{dy}{dt}\right) = X(x, y) = (x, y)$  donde el campo vectorial es un campo radial (ver figura [A.1](#)) entonces el flujo de esta ecuación viene dado por  $\exp(tX)(x, y) = (xe^t, ye^t)$ , luego  $X$  es exactamente integrable.

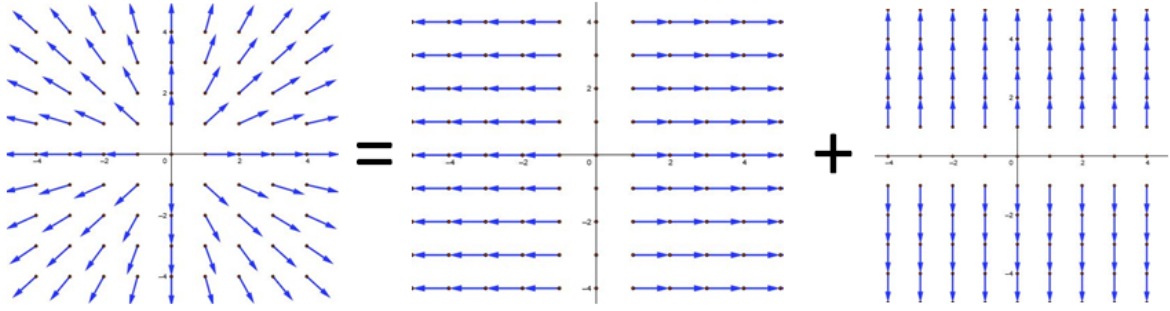


Figura A.1: Descomposición del campo vectorial  $X = (x, y)$  en los campos  $X^1 = (x, 0)$  y  $X^2 = (0, y)$ .

Considere el sistema  $\frac{dx}{dt} = X(x)$  en  $\mathbb{R}^n$  donde  $X$  no es exactamente integrable. Supongamos que podemos descomponer el campo vectorial como  $\frac{dx}{dt} = X^1(x) + X^2(x)$  de tal forma que los nuevos campos sean exactamente integrables y calculamos los flujos  $\exp(tX^1)$  y  $\exp(tX^2)$  de los sistemas  $\frac{dx}{dt} = X^1(x)$  y  $\frac{dx}{dt} = X^2(x)$  respectivamente.

La siguiente proposición se demuestra utilizando la fórmula de Baker-Campbell-Hausdorff y nos permite construir integradores de segundo orden (ver [Y90]).

**Proposición A.0.7.** *Sea  $X(x)$  un campo vectorial, si  $X$  se puede descomponer como una suma  $X(x) = \sum_{i=1}^n Y_i(x)$  donde cada uno de los campos  $Y_i(x)$  es exactamente integrable, entonces*

$$S_2(\tau) = \exp\left(\frac{\tau}{2}Y_1\right) \exp\left(\frac{\tau}{2}Y_2\right) \cdots \exp(\tau Y_n) \cdots \exp\left(\frac{\tau}{2}Y_2\right) \exp\left(\frac{\tau}{2}Y_1\right),$$

es un integrador de segundo orden para la ecuación diferencial  $\dot{x} = X(x)$ .

La proposición nos dice que si descomponemos el campo vectorial  $X$  en suma de campos vectoriales  $Y_i$  de tal forma que los campos  $Y_i$  pueden ser integrados de manera exacta entonces podemos combinar estas soluciones para tener un integrador de segundo orden para  $X$ .

Existen diferentes algoritmos para construir explícitamente integradores de orden par, que pueden consultarse en [BSVZ19, Proposición 7 y 8], los cuales nos permiten construir integradores de contacto (o simplécticos), es decir, que preserven la estructura de contacto (o simpléctica).

**Proposición A.0.8.** *Sea  $H$  un Hamiltoniano de contacto o simpléctico tal que  $H = \sum_{i=1}^n H_i$  y además cada campo  $X_{H_i}$  es exactamente integrable. Entonces el integrador*

$$S_2(\tau) = \exp\left(\frac{\tau}{2}X_{H_1}\right) \exp\left(\frac{\tau}{2}X_{H_2}\right) \cdots \exp(\tau X_{H_n}) \cdots \exp\left(\frac{\tau}{2}X_{H_2}\right) \exp\left(\frac{\tau}{2}X_{H_1}\right), \quad (\text{A.6})$$

es de segundo orden y preserva la estructura de contacto o simpléctica.

Como  $S_2$  es un integrador simpléctico o de contacto, entonces tiene un Hamiltoniano asociado, conocido como **Hamiltoniano sombra**, el cual se puede escribir como una serie usando la fórmula de Baker-Campbell-Hausdorff. Este Hamiltoniano sombra es importante porque hace que casi se preserve el Hamiltoniano original. Por otro lado, si en (A.6)



usamos la fórmula de Baker-Campbell-Hausdorff y tenemos en cuenta que  $X_H = \sum_{i=1}^n X_{H_i}$  se obtiene

$$S_2(\tau) = \exp(\tau X_H) + \mathcal{O}(\tau^3).$$

Tome  $N = \frac{T}{\tau}$ , entonces se puede aproximar la evolución de la dinámica para un tiempo  $T$  con error cuadrático, pues componiendo  $N$  veces se tiene

$$(S_2(\tau))^N = (\exp(\tau X_H))^N + \frac{T}{\tau} \mathcal{O}(\tau^3) = \exp(T X_H) + \mathcal{O}(\tau^2).$$

**Ejemplo A.0.9.** Considere el Hamiltoniano del oscilador armónico adimensional  $H = \frac{q^2+p^2}{2}$  y considere este como la suma de dos Hamiltonianos  $H_1 = \frac{q^2}{2}$  y  $H_2 = \frac{p^2}{2}$  así los respectivos campos Hamiltonianos son

$$\begin{aligned} X_{H_1} &= -q \frac{\partial}{\partial p} \\ X_{H_2} &= p \frac{\partial}{\partial q}. \end{aligned}$$

El integrador de segundo orden está dado por

$$S_2(\tau) = \exp\left(\frac{\tau}{2} X_{H_1}\right) \exp(\tau X_{H_2}) \exp\left(\frac{\tau}{2} X_{H_1}\right),$$

así solucionando las ecuaciones de flujo se tiene:

$$\begin{aligned} \exp\left(\frac{\tau}{2} X_{H_1}\right) \begin{bmatrix} q \\ p \end{bmatrix} &= \begin{bmatrix} q \\ p - \frac{q\tau}{2} \end{bmatrix} \\ \exp(\tau X_{H_2}) \begin{bmatrix} q \\ p \end{bmatrix} &= \begin{bmatrix} q + p\tau \\ p \end{bmatrix}. \end{aligned}$$

Realizando la composición se recuperan las ecuaciones (A.4) y (A.5).

**Ejemplo A.0.10.** Consideremos el Hamiltoniano para el oscilador armónico amortiguado.

$$\mathcal{H} = \frac{p^2}{2m} + \frac{1}{2}m\omega^2 q^2 + \gamma S.$$

Donde  $m$  es la masa de la partícula,  $\omega$  es la frecuencia angular y  $\gamma$  es el coeficiente de fricción. Tome  $H_1 = \frac{p^2}{2m}$ ,  $H_2 = \frac{1}{2}m\omega^2 q^2$  y  $H_3 = \gamma S$ . Luego el integrador de segundo orden según la proposición A.0.8 es

$$S_2(\tau) = \exp\left(\frac{\tau}{2} X_{H_1}\right) \exp\left(\frac{\tau}{2} X_{H_2}\right) \exp(\tau X_{H_3}) \exp\left(\frac{\tau}{2} X_{H_2}\right) \exp\left(\frac{\tau}{2} X_{H_1}\right).$$

Entonces utilizando la ecuación (1.25) se encuentra que

$$\begin{cases} X_{H_1} = \frac{p^2}{2m} \frac{\partial}{\partial S} + \frac{p}{m} \frac{\partial}{\partial q} \\ X_{H_2} = -\frac{1}{2}m\omega^2 q^2 \frac{\partial}{\partial S} - m\omega^2 q \frac{\partial}{\partial p} \\ X_{H_3} = -\gamma S \frac{\partial}{\partial S} - \gamma p \frac{\partial}{\partial p}. \end{cases}$$

Solucionando las ecuaciones del flujo para cada Hamiltoniano se obtiene:

$$\begin{aligned} \exp\left(\frac{\tau}{2}X_{H_1}\right) \begin{bmatrix} q \\ p \\ S \end{bmatrix} &= \begin{bmatrix} q + \frac{p\tau}{2m} \\ p \\ S + \frac{p^2\tau}{4m} \end{bmatrix} \\ \exp\left(\frac{\tau}{2}X_{H_2}\right) \begin{bmatrix} q \\ p \\ S \end{bmatrix} &= \begin{bmatrix} q \\ p - \frac{m\omega^2 q\tau}{2} \\ S - \frac{m\omega^2 q^2\tau}{4} \end{bmatrix} \\ \exp(\tau X_{H_3}) \begin{bmatrix} q \\ p \\ S \end{bmatrix} &= \begin{bmatrix} q \\ pe^{-\gamma\tau} \\ Se^{-\gamma\tau} \end{bmatrix}. \end{aligned}$$

Cuando se tiene un integrador geométrico, el siguiente teorema da el orden de error que hay entre un método de optimización discreto y uno continuo para cierto sistema Hamiltoniano. Este es el teorema principal en [\[FJV20\]](#).

**Teorema A.0.11.** *Considere un sistema Hamiltoniano disipativo dependiente del tiempo dado por [\(1.17\)](#), con función Hamiltoniana*

$$H = e^{-\eta_1(t)}T(q, p, t) + e^{\eta_2(t)}f(q),$$

con  $\eta_1$  y  $\eta_2$  funciones monótonas crecientes que son las que se encargan de dar disipación al sistema, además,  $T$  es la energía cinética y la consideramos Lipschitz. Sea  $\Phi_\tau$  un integrador de contacto con respecto a la estructura de contacto definida por la 1-forma de Poincaré-Cartan de orden  $r$ , tal que cumple la condición de Lipschitz

$$\|\Phi_\tau(y) - \Phi_\tau(x)\| \leq (1 + \tau L_\Phi)\|y - x\|,$$

para alguna constante  $L_\Phi > 0$  y para todo  $x, y \in T^*M$ . Entonces  $\Phi_\tau$  preserva la razón continua de convergencia salvo por un error pequeño dado por

$$f(q_k) - f(q^*) = f(q(t_k)) - f(q^*) + \mathcal{O}(\tau^r e^{-\eta_2(t_k)}) \quad (\text{A.7})$$

donde  $e^{L_\Phi t_k - \eta_1(t_k)} < \infty$  para  $t_k$  muy grande. Esto se tiene para tiempos grandes dados por  $t_k = \tau k = \mathcal{O}(\tau^r e^r e^{\frac{\tau_0}{\tau}})$  con  $\tau_0 > 0$  independiente de  $\tau$ .