

UNIVERSIDADE FEDERAL DE GOIÁS
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

TIAGO MOREIRA VARGAS

**Geometria Riemann-Finsler com
Aplicações em Geometria da
Informação**

Goiânia
2009

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



Termo de Ciência e de Autorização para Disponibilizar as Teses e Dissertações Eletrônicas (TEDE) na Biblioteca Digital da UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás-UFG a disponibilizar gratuitamente através da Biblioteca Digital de Teses e Dissertações - BDTD/UFG, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou download, a título de divulgação da produção científica brasileira, a partir desta data.

1. Identificação do material bibliográfico: Dissertação Tese

2. Identificação da Tese ou Dissertação

Autor(a):	TIAGO MOREIRA VARGAS		
CPF:		E-mail:	tiagomvargas@hotmail.com
Seu e-mail pode ser disponibilizado na página? <input checked="" type="checkbox"/> Sim <input type="checkbox"/> Não			
Vínculo Empregatício do autor			
Agência de fomento:	Conselho Nacional de Desenvolvimento Científico e Tecnológico	Sigla:	CNPQ
País:	Brasil	UF:	GO
CNPJ:			
Título:	GEOMETRIA Riemann - Finsler com aplicação em GEOMETRIA da Informação		
Palavras-chave:	Geometria de Finsler, Geometria da Informação		
Título em outra língua:	Riemann - Finsler Geometry with applications in Information Geometry		
Palavras-chave em outra língua:	Finsler Geometry, Information Geom		
Área de concentração:	Geometria e Topologia		
Data defesa: (dd/mm/aa)	20/01/09		
Programa de Pós-Graduação:	Mestrado em Matemática		
Orientador(a):	Prof. Marcelo Almeida de Souza		
CPF:		E-mail:	msouza@mat.ufg.br
Co-orientador(a):			
CPF:		E-mail:	

3. Informações de acesso ao documento:

Liberação para disponibilização?¹ total parcial

Em caso de disponibilização parcial, assinale as permissões:

Capítulos. Especifique: _____

Outras restrições: _____

Havendo concordância com a disponibilização eletrônica, torna-se imprescindível o envio do(s) arquivo(s) em formato digital PDF ou DOC da tese ou dissertação. O Sistema da Biblioteca Digital de Teses e Dissertações garante aos autores, que os arquivos contendo eletronicamente as teses e ou dissertações, antes de sua disponibilização, receberão procedimentos de segurança, criptografia (para não permitir cópia e extração de conteúdo, permitindo apenas impressão fraca) usando o padrão do Acrobat.

 Tiago Moreira Vargas
Assinatura do(a) autor(a)

Data: 02 / 02 / 2009

¹ Em caso de restrição, esta poderá ser mantida por até um ano a partir da data de defesa. A extensão deste prazo suscita justificativa junto à coordenação do curso. Todo resumo e metadados ficarão sempre disponibilizados.

TIAGO MOREIRA VARGAS

Geometria Riemann-Finsler com Aplicações em Geometria da Informação

Dissertação apresentada ao Programa de Pós-Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de concentração: Geometria e Topologia.

Orientador: Prof. Dr. Marcelo Almeida de Souza

Goiânia
2009

Dados Internacionais de Catalogação na Publicação (CIP)
(GPT/BC/UFG)

V297g Vargas, Tiago Moreira.
Geometria Riemann-Finsler com aplicações em geometria da
Informação [manuscrito] / Tiago Moreira Vargas. – 2009.
77f.

Orientador: Prof. Dr. Marcelo Almeida de Souza.

Dissertação (Mestrado) – Universidade Federal de Goiás,
Instituto de Matemática e Estatística, 2009.

Bibliografia: f.76-77.

Apêndice.

1. Geometria diferencial 2. Geometria riemanniana 3. Finsler,
geometria de 4. Geometria da informação I. Souza, Marcelo Almei-
da de II. Universidade Federal de Goiás, **Instituto de Matemática e**
Estatística. III. Título.

CDU: 514.7

TIAGO MOREIRA VARGAS

**“Geometria Riemann-Finsler com Aplicações em
Geometria da Informação”**

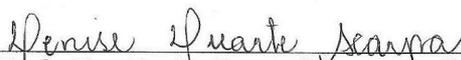
Dissertação defendida no Programa de Pós-Graduação do Instituto de Matemática e Estatística da Universidade Federal de Goiás como requisito parcial para obtenção do título de Mestre em Matemática, aprovada em 20 de janeiro de 2009, pela Banca Examinadora constituída pelos professores:



Prof. Dr. Marcelo Almeida de Souza
Instituto de Matemática e Estatística-UFG
Presidente da Banca



Prof. Dr. Armando Mauro Vasquez Corro
Instituto de Matemática e Estatística-UFG



Profa. Dra. Denise Duarte Scarpa Magalhães Alves
Instituto de Ciências Exatas-UFG

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador(a).

Tiago Moreira Vargas

Graduou-se Bacharel em Matemática pela Universidade Federal de Goiás (2006). Durante a graduação foi bolsista de Iniciação Científica do CNPq trabalhando nas áreas de Probabilidade e posteriormente em Geometria Diferencial. Durante o Mestrado, foi bolsista do CNPq.

Dedico este trabalho ao meu irmão
Hélio Vinícius Moreira Vargas (in memoriam).

Agradecimentos

Ao Nosso Senhor Jesus Cristo.

Aos meus pais: Hélio e Adelta.

Ao professor Marcelo, meu orientador, pela paciência e pelo tempo dedicado à minha orientação.

À todos os professores, funcionários e amigos do Instituto de Matemática e Estatística da UFG, em especial aos colegas Kamilla e Hugo.

Ao CNPq, pela ajuda financeira.

A todos que não citei, mas que de alguma forma contribuíram para a conclusão de mais uma etapa em minha vida.

O Barquinho

Dia de luz, festa de sol
E o barquinho a deslizar
No macio azul do mar
Tudo é verão, o amor se faz
Num barquinho pelo mar
Desliza sem parar...
Sem intenção, nossa canção
Vai saindo desse mar e o sol
Beija o barco e luz
Dias tão azuis
Beija o barco e luz
Dias tão azuis
Volta do mar, desmaia o sol
E o barquinho a deslizar
E a vontade é de cantar
Céu tão azul, ilhas do sul
O barquinho é o coração
Deslizando na canção
Tudo isso é paz, tudo isso traz
Uma calma de verão
E então
O barquinho vai, a tardinha cai
O barquinho vai, a tardinha cai

Ronaldo Bôscoli e Roberto Menescal,
1961 .

Resumo

Vargas, T.M.. **Geometria Riemann-Finsler com Aplicações em Geometria da Informação**. Goiânia, 2009. 78p. Dissertação de Mestrado. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

A Geometria da Informação é um novo ramo da matemática que se originou da aplicação da Geometria Diferencial à Estatística. Essa área surgiu a partir da investigação da estrutura geométrica de uma família de distribuições de probabilidade e tem sido aplicada com sucesso em diversas áreas como a Inferência Estatística e a Teoria da Informação. Essa dissertação é uma introdução à Geometria da Informação para um ponto de vista mais geral, utilizando a Geometria de Riemann-Finsler e a Geometria de Sprays.

Palavras-chave

Geometria de Riemann-Finsler e Geometria da Informação

Abstract

Vargas, T.M.. **Riemann-Finsler Geometry with applications in Information Geometry**. Goiânia, 2009. 78p. MSc. Dissertation. Instituto de Matemática e Estatística, Universidade Federal de Goiás.

Information Geometry is a new branch in mathematics, originated from the applications of Differential Geometry to Statistics. This area emerged from investigating the geometrical structure of a family of probability distributions, and has been applied to various areas including statistical inference and information theory. This dissertation is an introduction to the geometry of information for a more general point of view using Riemann-Finsler Geometry and Spray Geometry.

Keywords

Riemann-Finsler Geometry and Information Geometry

Sumário

1	Introdução	10
2	Preliminares	12
2.1	Espaços Riemannianos	12
2.2	Métricas de Finsler	13
2.3	H -função	16
2.4	Sprays	17
2.4.1	Sprays de Métricas Finsler	17
2.4.2	Sprays em Variedades	19
2.5	O Teorema de Euler para funções homogêneas	20
2.6	Probabilidade: Algumas definições	20
3	Divergências e Estruturas de Informação	24
3.1	Divergência: Definição.	24
3.2	Divergências Regulares	25
3.3	Estruturas de Informação	28
3.4	α -sprays de uma Estrutura de Informação	28
3.5	Métricas Finsler Dualmente Flat	37
3.6	Divergências afins e Estruturas de Informação afins	41
3.7	Estruturas de Informação afins do tipo α -Flat	43
3.8	Conexões Afins Dualísticas	49
4	Aplicações aos Modelos estatísticos	54
4.1	f -divergências em espaços de probabilidade	54
4.2	Os α -sprays em Modelos Estatísticos	57
4.3	Família exponencial de distribuições de probabilidade	67
4.4	Dualidade das f -divergências em espaços de probabilidade	73
	Referências Bibliográficas	76
	Referências Bibliográficas	76
A	Notações	78

Introdução

A Geometria da Informação surgiu para investigar a estrutura geométrica de famílias paramétricas de distribuições de probabilidade e tem sido aplicada na inferência estatística, teoria dos sistemas de controle, psicologia matemática, etc. Um exemplo seria, se considerarmos o conjunto S das distribuições normais com média μ e variância σ^2 :

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

S pode ser vista como uma variedade de dimensão 2, onde (μ, σ) é um sistema de coordenadas. No entanto, este não é um espaço euclidiano, mas sim um espaço com uma métrica Riemanniana que, naturalmente, decorre das propriedades das distribuições de probabilidade subjacentes. Em particular, S além de uma família de distribuições normais, é um espaço de curvatura constante negativa.

Nessa dissertação, iremos dar um enfoque mais geral à geometria da informação, utilizando para isto, a Geometria de Riemann-Finsler.

S.S.Chern disse: *A geometria de Finsler é exatamente a geometria Riemanniana sem a restrição quadrática.*.[18]

Em 1854, Riemann introduziu uma estrutura métrica em um espaço geral, que se baseou no elemento de arco $ds = L(x^i, dx^i)$, $1 \leq i \leq n$. Na geometria de Finsler, $L(x, y)$ é uma função positiva para $y \neq 0$ no fibrado tangente TM , e é homogênea de grau 2 em y .

Em 1918, Paul Finsler estudou os espaços métricos mais gerais, hoje chamados espaços Riemann-Finsler ou simplesmente espaços de Finsler.

Essa dissertação é baseada no artigo "*Riemann Finsler Geometry with Applications to Information Geometry*" de autoria de Z.Shen, publicado em 2006 por *Chinese Annals of Mathematics*. [4].

No capítulo 2, apresentaremos alguns resultados e conceitos de Geometria de Riemann-Finsler, e também alguns conceitos de probabilidade, que foram de suma importância para o desenvolvimento dessa dissertação.

No capítulo 3, introduziremos o conceito de Divergência, Divergências Regulares e Estruturas de Informação. Definiremos o que vem a ser um α -spray de uma Es-

trutura de informação, e provaremos alguns resultados. Definiremos Métricas de Finsler Dualmente Flat, Divergências Afins e Estruturas de informação afim α -flat, provando alguns resultados.

No capítulo 4, aplicaremos a teoria desenvolvida no capítulo 3 para o espaço \mathcal{P} das distribuições de probabilidade. Definiremos f -divergências em espaços de probabilidade, definiremos α -sprays em modelos estatísticos e aplicaremos as teorias desenvolvidas para as famílias exponenciais de distribuições de probabilidade. Falaremos brevemente sobre dualidade de f -divergências.

Abaixo estão relacionados alguns dos principais teoremas do trabalho, embora ele seja composto de vários outros resultados.

Teorema 1.1 *Seja $\alpha \neq 0$. Uma estrutura de informação (L, H) afim é α -flat se e somente se sua estrutura de informação dual (L^*, H^*) é α -flat.*

E também

Teorema 1.2 *Seja $f = f(t)$ uma função $f(1) = 0$ e $f''(1) = 1$. Para todo modelo estatístico regular (M, D_f) de $(\mathcal{P}, \mathcal{D}_f)$, a estrutura de informação induzida em M é dada por $(L_f, H_f) = (L, \rho N)$, onde $\rho := 3 + f'''(1)$ e*

$$L = \int_X \left[y^i \frac{\partial}{\partial x^i} \ln p \right]^2 p dr = \mathbb{E} \left[\left(\frac{\partial \ln p(r; x)}{\partial x^i} y^i \right)^2 \right], \quad (1-1)$$

$$N = \frac{1}{6} \int_X \left[y^i \frac{\partial}{\partial x^i} \ln p \right]^3 p dr = \mathbb{E} \left[\frac{1}{6} \left(\frac{\partial \ln p(r; x)}{\partial x^i} y^i \right)^3 \right], \quad (1-2)$$

onde \mathbb{E} denota a esperança matemática. O α -spray $G_{\alpha, \rho}$ de D_f é dado por

$$G_{\alpha, \rho}^i = \bar{G}^i + (\rho\alpha + 1)A^i,$$

onde

$$\bar{G}^i = \frac{g^{il}(x)}{2} \int_X \left[y^j y^j \frac{\partial^2}{\partial x^i \partial x^j} \ln p(r; x) \right] \frac{\partial}{\partial x^i} p(r; x) dr \quad (1-3)$$

e

$$A^i = \frac{g^{il}(x)}{4} \int_X \left[y^j \frac{\partial}{\partial x^j} \ln p(r; x) \right]^2 \frac{\partial}{\partial x^i} p(r; x) dr. \quad (1-4)$$

Preliminares

2.1 Espaços Riemannianos

Nessa seção, iremos apresentar alguns conceitos de Geometria Riemanniana, que serão utilizados nesse trabalho.

Definição 2.1 Uma Variedade Diferenciável de dimensão n é um conjunto M e uma família de aplicações biunívocas $x_\alpha : U_\alpha \subset \mathbb{R}^n \rightarrow M$ de abertos U_α de \mathbb{R}^n em M tais que

$$(i) \bigcup_{\alpha} x_\alpha(U_\alpha) = M.$$

(ii) Para todo par α, β , com $x_\alpha(U_\alpha) \cap x_\beta(U_\beta) = W \neq \emptyset$, os conjuntos $x_\alpha^{-1}(W)$ e $x_\beta^{-1}(W)$ são abertos em \mathbb{R}^n e as aplicações $x_\beta^{-1} \circ x_\alpha$ são diferenciáveis.

(iii) A família $\{(U_\alpha, x_\alpha)\}$ é máxima relativamente às condições (i) e (ii).

Observação 1 O par (U_α, x_α) com $p \in x_\alpha(U_\alpha)$ é chamado de **parametrização** ou **sistema de coordenadas** de M em p e $x_\alpha(U_\alpha)$ é chamada de **vizinhança coordenada** em p . Uma família $\{(U_\alpha, x_\alpha)\}$ satisfazendo as condições (i) e (ii) chama-se **estrutura diferenciável** em M .

Seja $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathbb{R}^n$ uma curva diferenciável em \mathbb{R}^n , com $\gamma(0) = p$ e $\gamma'(0) = v \in \mathbb{R}^n$. Seja f uma função diferenciável definida em uma vizinhança de p . Restringindo f à curva γ , podemos escrever a derivada direcional segundo o vetor $v \in \mathbb{R}^n$ como

$$\frac{d(f \circ \gamma)}{dt} \Big|_{t=0} = \sum_{i=1}^n \frac{df}{dx_i} \Big|_{t=0} \frac{dx_i}{dt} \Big|_{t=0} = \left(\sum_i x'_i(0) \frac{\partial}{\partial x_i} \right) f.$$

Logo, a derivada direcional segundo v é um operador sobre as funções diferenciáveis que depende somente de v .

Definição 2.2 Seja M uma variedade diferenciável. Uma aplicação diferenciável $\gamma : (-\varepsilon, \varepsilon) \rightarrow M$ é chamada de **curva diferenciável** em M . Suponha que $\gamma(0) = p \in M$, e

seja S o conjunto das funções de M diferenciáveis em p . O vetor tangente à curva γ em $t = 0$ é uma função $\dot{\gamma}(0) : S \rightarrow \mathbb{R}$ dada por

$$\dot{\gamma}(0)f = \frac{d(f \circ \gamma)}{dt} \Big|_{t=0}, f \in S.$$

Um vetor tangente em p é o vetor tangente em $t = 0$ de alguma curva γ em M com $\gamma(0) = p$. O conjunto dos vetores tangentes a M em p é indicado por T_pM .

Definição 2.3 Sejam M uma variedade diferenciável e seja $TM = \{(p, v) : p \in M, v \in T_pM\}$. O conjunto TM munido de uma estrutura diferenciável; será chamado de **fibrado tangente** de M .

Definição 2.4 Um **campo de vetores** X em uma variedade diferenciável M é uma correspondência que associa cada ponto p de M a um vetor $X(p) \in T_pM$. Visualizando como aplicação, X é uma aplicação de M no fibrado tangente TM . O campo de vetores X é diferenciável se a aplicação $X : M \rightarrow TM$ é diferenciável.

Definição 2.5 Uma **métrica Riemanniana** em uma variedade diferenciável M é uma correspondência que associa a cada ponto $p \in M$ um produto interno $\langle \cdot, \cdot \rangle_p$, isto é, uma forma bilinear, simétrica e positiva definida no espaço tangente T_pM , que varia diferenciavelmente no seguinte sentido: Se $x : U \subset \mathbb{R}^n \rightarrow M$ é um sistema de coordenadas locais em torno de p , com $x(x_1, \dots, x_n) = q \in x(U)$ e $\frac{\partial}{\partial x_i}(q) = dx_q e_i$, onde e_i é um vetor da base canônica de \mathbb{R}^n , então $\left\langle \frac{\partial}{\partial x_i}(q), \frac{\partial}{\partial x_j}(q) \right\rangle_q = g_{ij}(x_1, \dots, x_n)$ é uma função diferenciável em U .

2.2 Métricas de Finsler

Seja M uma variedade C^∞ n -dimensional e $\pi : TM \rightarrow M$ a projeção natural. Denotemos por (x, y) um ponto de TM , $x \in M$, $y \in T_xM$. Consideremos $\widetilde{TM} := TM \setminus \{0\}$.

Introduzimos, a seguir, o conceito de métrica (estrutura) de Finsler.

Definição 2.6 Uma métrica de Finsler em M é uma função $L : TM \rightarrow [0, \infty)$ que satisfaz às seguintes condições:

- (i) (Regularidade) $L \in C^\infty$ em \widetilde{TM} ;
- (ii) (Homogeneidade positiva) $L(x, \lambda y) = \lambda^2 L(x, y)$, $\forall \lambda > 0$, $(x, y) \in TM$;
- (iii) (Convexidade forte) $g = (g_{ij}(x, y)) = \left(\frac{1}{2}[L(x, y)]_{y_i y_j}\right)$ é positiva definida em cada ponto de \widetilde{TM} .

Vamos esclarecer melhor o significado das coordenadas y^i 's no item [iii] da Definição 2.6. Consideremos uma base qualquer $\{e_i\}$ para T_xM . Podemos tomar $e_i = \frac{\partial}{\partial x^i}$, apesar de que esta restrição é desnecessária. Expressando y como $y^i e_i$, a estrutura de Finsler L é então uma função de (x^i, y^i) , e as derivadas parciais de $\frac{1}{2}L$ são tomadas em relação aos y^i 's. Pode ser mostrado que o fato de g , dada no item [iii], ser positiva definida independe de nossa escolha para $\{e_i\}$.

Definição 2.7 *Um espaço (M, L) é dito uma variedade de Finsler.*

Faremos daqui para frente o uso da convenção de Einstein, ou seja, não escreveremos o símbolo do somatório para representar a soma quando aparecerem índices repetidos. Salvo em menção contrária, onde explicitamente colocaremos o símbolo do somatório. No que segue, com relação ao sistema de coordenadas locais escolhido vamos considerar o seguinte: Seja $(x^1, \dots, x^n) = (x^i) : U \rightarrow \mathbb{R}^n$ um sistema de coordenadas locais sobre um aberto $U \subset M^n$. Como é usual, $\{\frac{\partial}{\partial x^i}\}$ e $\{dx^i\}$ são respectivamente as bases de coordenadas induzidas para T_xM e T_x^*M . Os x^i 's dão origem às coordenadas locais (x^i, y^i) sobre $\pi^{-1}(U) \subset TM$ por meio de

$$y = y^i \frac{\partial}{\partial x^i}.$$

As funções L que são definidas sobre TM podem ser localmente expressas como

$$L(x^1, \dots, x^n, y^1, \dots, y^n),$$

e as derivadas parciais serão denotadas por $L_{x^i}, L_{x^i x^j}, \dots, L_{y^i}, L_{y^i y^j}, \dots$.

Exemplo 2.8 *Uma função métrica F de Finsler será Riemanniana se $L(x, y) = a_{ij}(x)y^i y^j$. Nesse caso, $g_{ij}(x, y) = \frac{1}{2}[L]_{y^i y^j} = a_{ij}(x)$ independe de y , para cada i, j . Temos ainda que esta métrica será Euclidiana se os g_{ij} 's independem de x também.*

Exemplo 2.9 *Uma métrica de Finsler L em um espaço vetorial real n -dimensional V^n que independe de $x \in V^n$ é dita do tipo Minkowski. A classe das variedades de Minkowski é a mais simples das variedades Finslerianas. Temos ainda que para cada $x \in M$ fixado, o espaço (T_xM, L_x) é do tipo Minkowski. Um exemplo não-Riemanniano nem do tipo Minkowski é dado por*

$$L(x, y) = \sum_{i=1}^n (y^i)^2 + \lambda(x) \left[\sum_{k=1}^n |y^k|^{2m} \right]^{\frac{1}{m}},$$

onde $\lambda : M \rightarrow \mathbb{R}$ é uma função positiva diferenciável.

Exemplo 2.10 Outro importante exemplo será dado na próxima definição.

G. Randers, em 1941, estudou uma classe de métricas de Finsler, a qual leva o seu nome. Para introduzirmos o conceito de espaço de Randers precisamos de duas estruturas,

- uma métrica Riemanniana $ds^2 := a_{ij}dx^i \otimes dx^j$, sobre uma variedade M^n de classe C^∞ ,
- uma 1-forma controlada $\beta := b_k dx^k$ sobre M^n .

Juntamente elas definem de modo bem simples uma métrica de Finsler. Vejamos agora esta interessante classe de métricas.

Definição 2.11 Uma métrica de Randers em M é uma estrutura Finsler sobre TM dada por

$$L(x, y) = (\alpha(x, y) + \beta(x, y))^2, \quad (2-1)$$

onde

$$\alpha(x, y) = \sqrt{a_{ij}(x)y^i y^j}, \quad (2-2)$$

$$\beta(x, y) = b_k(x)y^k, \quad (2-3)$$

e a_{ij} , a^{ij} são as componentes da matriz de uma métrica riemanniana e de sua inversa, respectivamente, e b_k são as componentes da 1-forma β , cuja norma

$$b = \sqrt{a^{ij}b_i b_j}, \quad (2-4)$$

satisfaz $0 \leq b < 1$. Observemos que $b(x) = \sup_{y, \alpha(x, y)=1} \beta(x, y)$.

Vamos apresentar agora a noção de comprimento finsleriano de uma curva e depois veremos uma função distância entre dois pontos. Omitiremos as demonstrações e detalhes, os quais podem ser encontrados em [1].

Os conceitos de comprimento de Finsler de uma curva e de função distância de uma métrica de Finsler serão apresentado na seguinte definição.

Definição 2.12 Seja (M, L) uma variedade de Finsler, onde L é positivamente homogênea de grau 2. Seja $c : [0, 1] \rightarrow M$ uma curva diferenciável por partes com velocidade $\frac{dc}{dt} = \frac{dc^i}{dt} \frac{\partial}{\partial x^i} \in T_{c(t)}M$. O comprimento integral $L(c)$ é definido como

$$L(c) = \int_0^1 \sqrt{L\left(c(t), \frac{dc}{dt}\right)} dt. \quad (2-5)$$

Utilizando essa estrutura de comprimento, podemos definir uma função $d = d(p, q)$ em $M \times M$ por

$$d(p, q) = \inf(L(C)),$$

onde o ínfimo é calculado sobre todas as curvas que ligam p a q , com $c(0) = p$ e $c(1) = q$. A função distância d satisfaz

- (i) $d(p, q) \geq 0$ e a igualdade acontece se, e somente se, $p = q$;
- (ii) $d(p, q) \leq d(p, r) + d(r, q)$.

A função d é chamada de função distância de L .

2.3 H -função

Vamos agora introduzir o conceito de H -função, que será primordial nos capítulos subseqüentes desse trabalho.

Definição 2.13 Uma H -função em uma variedade M é uma função escalar $H = H(x, y)$ em TM com as seguintes propriedades:

- (i) $H(x, \lambda y) = \lambda^3 H(x, y)$, $\lambda > 0$.
- (ii) $H(x, y)$ é de classe C^∞ em \widetilde{TM} .

As H -funções são funções positivamente homogêneas de grau 3. Exemplos de H -função serão dados a seguir.

Exemplo 2.14 Se $L = L(x, y)$ é uma métrica de Finsler em uma variedade M , então a função

$$H(x, y) = L(x, y)^{\frac{3}{2}},$$

é uma H -função em M . De fato, L é de classe C^∞ em $TM \setminus \{0\}$, logo H também o é, e para $\lambda > 0$,

$$\begin{aligned} H(x, \lambda y) &= L(x, \lambda y)^{\frac{3}{2}} \\ &= (\lambda^2)^{\frac{3}{2}} L(x, y)^{\frac{3}{2}} \\ &= \lambda^3 H(x, y). \end{aligned}$$

Outro exemplo de H -função, que será muito utilizado nesse trabalho, será:

Exemplo 2.15 Para $L = L(x, y)$ uma métrica de Finsler em um subconjunto $\Omega \subset \mathbb{R}^n$, defina

$$H(x, y) = \frac{1}{2} L_{x^k}(x, y) y^k.$$

Temos que H é uma H -função em Ω , pois como L é de classe C^∞ em $TM \setminus \{0\}$, então H também o é, e para $\lambda > 0$,

$$\begin{aligned} H(x, \lambda y) &= \frac{1}{2} L_{x^k}(x, \lambda y) (\lambda y)^k \\ &= \frac{1}{2} \lambda^3 L_{x^k}(x, y) y^k \\ &= \lambda^3 H(x, y). \end{aligned}$$

2.4 Sprays

Nesta seção, iremos dar a definição do campo de vetores *spray*, tanto em variedades Finsler, como em variedades quaisquer. Essas definições serão necessárias para os cálculos que serão efetuados nos capítulos 2 e 3.

2.4.1 Sprays de Métricas Finsler

Definição 2.16 Um *spray* sobre uma variedade Finsler M munida de uma métrica L é um campo de vetores definidos sobre TM o qual é expresso num sistema padrão de coordenadas locais (x^i, y^i) por

$$\mathcal{G} = y - \mathcal{G}^i(x, y) \frac{\partial}{\partial y^i},$$

onde $y = y^i \frac{\partial}{\partial x^i}$ é um campo de vetores e os coeficientes $\mathcal{G}^i(x, y)$ são funções de classe C^∞ , definidas localmente sobre $TM \setminus \{0\}$ satisfazendo

$$\mathcal{G}^i(x, \lambda y) = \lambda^2 \mathcal{G}^i(x, y),$$

para $\lambda > 0$ com

$$\mathcal{G}^i(x, y) = \frac{1}{4} g^{il}(x, y) \{L_{x^k y^l}(x, y) y^k - L_{x^l}(x, y)\}. \quad (2-6)$$

Uma idéia sobre a origem da equação (2-6) está no fato de se utilizar um método variacional para se achar a equação das geodésicas de uma variedade Finsler. Primeiramente, sabemos que para uma variedade finsler (M, L) , o comprimento de uma curva C parametrizada por $c = c(t)$, $0 \leq t \leq 1$ é dado por (2-5). Vamos parametrizar as

curvas que ligam $c(0)$ a $c(1)$ da seguinte maneira:

$$c_s(t) = H(s, t),$$

$$c(t) = H(0, t), 0 \leq t \leq 1.$$

Vamos assumir que c tenha velocidade constante, ou seja,

$$L(c(t), \dot{c}(t)) = \text{const.}$$

Suponha que c tenha comprimento mínimo. Então $L'(0) = 0$ em qualquer variação. Temos então

$$L(s) = \int_0^1 \sqrt{L\left(H(s, t), \frac{\partial}{\partial t} H(s, t)\right)} dt,$$

e

$$L'(0) = - \int_0^1 \frac{1}{\sqrt{L}} g_{jk}(x, y) \left\{ \frac{d^2 c^j}{dt^2} + 2\mathcal{G}^j\left(c(t), \frac{dc}{dt}\right) \right\} v^k dt,$$

onde $v^k = V^i(t) \frac{\partial}{\partial x^i} = \frac{\partial H(0, t)}{\partial s}$ e

$$\mathcal{G}^j(x, y) = \frac{1}{4} g^{il}(x, y) \{L_{x^k y^l}(x, y) y^k - L_{x^l}(x, y)\}.$$

Fazendo $L'(0) = 0$ para todo $V(t)$, temos

$$\ddot{c}(t) + 2\mathcal{G}^i(c, \dot{c}(t)) = 0. \quad (2-7)$$

Essa é a equação das geodésicas do spray da métrica de Finsler L .

Observação 2 (1) *É possível existir duas métricas de Finsler distintas com o mesmo spray. Por exemplo, se L é uma métrica de Finsler arbitrária em uma variedade, então a métrica $\bar{L} = kL$ tem o mesmo Spray de L para alguma constante positiva k .*

(2) *Se $L = g_{ij}(x)y^i y^j$ é uma métrica Riemanniana, então*

$$\mathcal{G}^i(x, y) = \frac{1}{2} \Upsilon_{jk}^i(x) y^j y^k, \quad \Upsilon_{jk}^i(x) = \Upsilon_{kj}^i(x), \quad (2-8)$$

onde

$$\Upsilon_{jk}^i(x) = \frac{g^{il}(x)}{2} \left\{ \frac{\partial g_{il}(x)}{\partial x^k} + \frac{\partial g_{kl}(x)}{\partial x^j} - \frac{\partial g_{jk}(x)}{\partial x^l} \right\}. \quad (2-9)$$

As funções locais $\Upsilon_{jk}^i(x)$ são chamadas **Símbolos de Christoffel**. Notemos que \mathcal{G}^i é quadrático em y .

(3) *Uma métrica de Finsler L é chamada de **métrica de Berwald** se os coeficientes do spray, ou seja \mathcal{G}^i é quadrático em y . Existem muitas métricas de Berwald não-Riemannianas.*

2.4.2 Sprays em Variedades

O conceito de sprays induzidos por Métricas de Finsler pode ser generalizado.

Definição 2.17 *Seja M uma variedade. Um spray G em M é um campo de vetores no fibrado tangente TM tal que em algum sistema de coordenadas locais usual (x^i, y^i) em TM , é expressado da seguinte forma:*

$$G = y^i \frac{\partial}{\partial x^i} - 2G^i \frac{\partial}{\partial y^i}, \quad (2-10)$$

onde $G^i \in C^\infty$ com $y \neq 0$ e

$$G^i(x, \lambda y) = \lambda^2 G^i(x, y), \quad \lambda > 0.$$

Uma curva $\sigma = \sigma(t)$ é chama de *geodésica* de G em uma variedade M se satisfaz o seguinte sistema de equações:

$$\ddot{x}(t) + 2G^i(x, \dot{x}(t)) = 0,$$

onde $x(t) = (x^i(t))$ denota as coordenadas de $\sigma(t)$. Geodésicas também são chamadas de *caminhos*. A coleção de todos os caminhos de um spray é chamada *estrutura de caminhos*.

Um spray G é denominado *afim* se em algum sistema de coordenadas locais,

$$G^i(x, y) = \frac{\Gamma_{jk}^i(x)}{2} y^j y^k, \quad \Gamma_{jk}^i(x) = \Gamma_{kj}^i(x). \quad (2-11)$$

Por definição, uma métrica de Finsler é uma métrica de Berwald se e somente se os seus sprays são afins. Todo spray afim G com coeficientes dados por (2-11), define uma conexão ∇ em TM ,

$$\nabla_y X = \{dX^i(y) + X^j \Gamma_{jk}^i(x) y^k\} \frac{\partial}{\partial x^i} \Big|_x, \quad (2-12)$$

onde $X = X^i \frac{\partial}{\partial x^i} \in C^\infty(TM)$ e $y = y^i \frac{\partial}{\partial x^i} \Big|_x \in T_x M$. ∇ é linear no seguinte sentido:

$$\nabla_{\lambda y + \mu v} X = \lambda \nabla_y X + \mu \nabla_v X,$$

$$\nabla_y (X + Y) = \nabla_y X + \nabla_y Y,$$

$$\nabla_y (fX) = df_x(y)X + f(x) \nabla_y X,$$

onde $y, v \in T_x M$, $f \in C^\infty(M)$ e $X, Y \in C^\infty(TM)$. A conexão ∇ é *compatível com a métrica* ou é *livre de torção* se

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

onde $X, Y \in C^\infty(TM)$. Conexões com torção livre e lineares são chamadas de *conexões afins*. Todo spray afim define uma conexão afim por (2-12). Inversamente, toda conexão

afim ∇ em TM define um spray afim como em (2-11). Logo as conexões afins tem correspondência 1 – 1 com os sprays afins.

Definição 2.18 *Um spray G em uma variedade é chamado **flat** se para todo ponto, existe um sistema de coordenadas locais (x^i, y^i) em TM tal que $G = y^i \frac{\partial}{\partial x^i}$, isto é, $G^i = 0$. Nesse caso (x^i, y^i) é chamado de sistema de coordenadas adaptado.*

Os sprays do tipo *flat* são casos muito especiais de sprays afins. Se G é flat, então em um sistema de coordenadas locais, as geodésicas são linhas retas, isto é, as coordenadas $(x^i(t))$ de toda geodésica $\sigma(t)$ assumem a forma

$$\dot{x}^i(t) = a^i t + b^i, \quad \forall t.$$

2.5 O Teorema de Euler para funções homogêneas

A seguir, teremos o Teorema de Euler para funções homogêneas, que será muito utilizado nos capítulos 2 e 3 desse trabalho.

Teorema 2.19 *Seja A um subconjunto aberto do \mathbb{R}^m , tal que, se $y \in A$, então $ty \in A$ e seja $L : A \rightarrow \mathbb{R}$ uma função diferenciável. Então para $k \in \mathbb{R}$, L é homogênea de grau k , isto é, $L(ty) = t^k L(y)$ se, e somente se,*

$$y^i \frac{\partial L}{\partial y^i}(y) = kL(y), \quad \forall y \in A. \quad (2-13)$$

A demonstração desse teorema pode ser encontrada em [1]. Como consequência desse teorema, temos as algumas propriedades. Considere F uma função homogênea de grau 1 em \mathbb{R}^n . Então

- (i) $y^j F_{y^j y^j} = 0, \quad \forall i.$
- (ii) $y^k F_{y^i y^j y^k}(y) = -F_{y^i y^j}(y), \quad \forall i, j.$
- (iii) $y^l F_{y^i y^j y^k y^l}(y) = -2F_{y^i y^j y^k}(y) \quad \forall i, j, k.$

2.6 Probabilidade: Algumas definições

Nessa seção daremos alguns conceitos básicos de probabilidade, que serão utilizadas no decorrer do trabalho.

Definição 2.20 *Seja (Ω, \mathcal{A}) um espaço mensurável, onde \mathcal{A} é uma σ -álgebra. Uma medida de probabilidade em (Ω, \mathcal{A}) é uma aplicação $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ tal que*

$$(i) \mathbb{P}(\emptyset) = 0;$$

$$(ii) \mathbb{P}(\Omega) = 1;$$

$$(iii) \text{ se } A_n, n \geq 1 \text{ é uma seqüência disjunta, então } \mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

A terna $(\Omega, \mathcal{A}, \mathbb{P})$ é chamada de *espaço de probabilidade*. Cada elemento de \mathcal{A} é chamado de *evento*. Chamamos de $\mathbb{P}(A)$, $A \in \mathcal{A}$, a probabilidade do evento A ocorrer. Maiores detalhes e propriedades podem ser vistas em [15].

Definição 2.21 Uma *variável aleatória* X em um espaço de probabilidade $(\Omega, \mathcal{A}, \mathbb{P})$ é uma função real definida no espaço Ω tal que $\{X \leq x\}$ é evento aleatório para todo $x \in \mathbb{R}$, isto é, $X : \Omega \rightarrow \mathbb{R}$ é variável aleatória se $\{X \leq x\} \in \mathcal{A}, \forall x \in \mathbb{R}$.

Observação 3 Em Teoria da Medida, $\{X \leq x\} \in \mathcal{A}, \forall x \in \mathbb{R}$ significa que X é uma função mensurável em \mathcal{A} , ou seja, se considerarmos (Ω, \mathcal{A}) e $(\overline{\Omega}, \overline{\mathcal{A}})$ espaços mensuráveis, então para $X : \Omega \rightarrow \overline{\Omega}$ (nesse caso $\overline{\Omega} = \mathbb{R}$), temos que $\forall \overline{A} \in \overline{\mathcal{A}}, X^{-1}(\overline{A}) \in \mathcal{A}$. Maiores detalhes sobre espaços mensuráveis, ver em [16].

Definição 2.22 A função de distribuição da variável aleatória X , denotada por F_X ou simplesmente por F , é definida por

$$F_X(x) = \mathbb{P}(X \leq x), \quad x \in \mathbb{R}.$$

Observação 4 Podemos chamar a função de distribuição de X de função de distribuição acumulada de X .

Para maiores detalhes sobre propriedades de variáveis aleatórias, ver em [15]. As variáveis aleatórias podem ser *discretas* ou *contínuas*. Uma variável aleatória X é discreta se ela toma um número finito ou pelo menos enumerável de valores. Nesse caso, chamamos a função $\mathbb{P}(x_i) = \mathbb{P}(X = x_i), i = 1, \dots$ de *função de probabilidade* ou *função de frequência* de X . Uma variável aleatória X é contínua se existe uma função $f(x) \geq 0$ tal que

$$F_X(x) = \int_{-\infty}^x f(t) dt, \forall x \in \mathbb{R}.$$

Neste caso, chamamos f de *função densidade de probabilidade* de X ou *densidade* de X .

Definição 2.23 Seja X uma variável aleatória contínua qualquer e F sua função de distribuição. A *esperança* de X é definida por

$$EX = \int_{-\infty}^{\infty} x dF(x) = \int_{-\infty}^{\infty} x f(x) dx, \quad (2-14)$$

onde f é a densidade de X .

Maiores detalhes sobre esperança de variáveis aleatórias discretas e propriedades da esperança, podem ser vistas em [15]. Uma definição mais rigorosa, do ponto de vista da Teoria medida pode ser vista em [16].

Definição 2.24 *Se $EX < \infty$, dizemos que X é integrável.*

Iremos agora dar alguns exemplos de variáveis aleatórias contínuas e integráveis.

Exemplo 2.25 (*Normal Gaussiana*) *Uma variável aleatória contínua X tem densidade Normal Gaussiana com esperança μ e variância σ^2 se sua função densidade de probabilidade é dada por*

$$n(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2-15)$$

onde $\mu \in \mathbb{R}$ e $\sigma > 0$.

Geometricamente, se temos os parâmetros μ e σ variando no conjunto $U = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma > 0\}$, então a família de distribuições normais $\{n(\mu, \sigma^2)\}$ é uma variedade de dimensão 2 em \mathbb{R}^3 . Através de um campo de vetores spray podemos determinar suas geodésicas. Outra maneira de calcular suas geodésicas pode ser vista em [5]. Seja X uma variável aleatória com densidade normal $n(0, \sigma^2)$. Temos então a seguinte densidade:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}, -\infty < x < \infty.$$

De acordo com esta fórmula, temos que $Y = X^2$ tem uma densidade \bar{f} dada por $\bar{f}(y) = 0$ para $y \leq 0$, e podemos ver facilmente que

$$\begin{aligned} \bar{F}_Y &= \mathbb{P}(X^2 \leq y) \\ &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f(t) dt. \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}). \end{aligned}$$

Diferenciando o último termo da equação em relação a y , temos

$$\bar{f}(y) = \frac{1}{2\sqrt{y}} (f(-\sqrt{y}) + f(\sqrt{y})), y > 0.$$

Logo

$$\bar{f}(y) = \frac{1}{\sigma\sqrt{2\pi y}} e^{-\frac{y}{2\sigma^2}}, y > 0. \quad (2-16)$$

Consideremos as funções $\bar{f}(x)$ da forma

$$\bar{f}(x) = \begin{cases} x^{\alpha-1} e^{-\lambda x} & , x > 0 \\ 0 & , \text{caso contrário.} \end{cases}$$

Verifica-se facilmente que se $\alpha > 0$ e $\lambda > 0$, a função \bar{f} é integrável. A densidade dada por (1-15) é o caso em que $\alpha = \frac{1}{2}$ e $\lambda = \frac{1}{2\sigma^2}$. Na normalização de \bar{f} para transformá-lo em densidade, devemos ter

$$\int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx = c.$$

Fazendo a mudança de variáveis $x = \frac{y}{\lambda}$ temos

$$c = \frac{1}{\lambda^{\alpha}} \int_0^{\infty} y^{\alpha-1} e^{-y} dy.$$

Como não podemos achar uma primitiva simples para a função acima, definimos então

$$c = \frac{1}{\lambda^{\alpha}} \Gamma(\alpha),$$

onde

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy.$$

A partir dessas observações, podemos dar a seguinte definição:

Definição 2.26 Uma função normalizada chama-se **densidade gama** de parâmetros α e λ , representada por $\Gamma(x; \alpha, \lambda)$ se tem a seguinte forma:

$$\Gamma(x; \alpha, \lambda) = \begin{cases} \frac{\lambda^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} & , x > 0 \\ 0 & , \text{caso contrário.} \end{cases}$$

onde

$$\Gamma(\alpha) = \lambda^{\alpha} \int_0^{\infty} x^{\alpha-1} e^{-\lambda x} dx.$$

Propriedades sobre a função $\Gamma(\alpha)$ podem ser vistas em [17].

Divergências e Estruturas de Informação

Neste capítulo iremos introduzir os conceitos de Divergências e Divergências regulares. Esses conceitos serão utilizados juntamente com as propriedades da H -função e também de sprays de métricas de Finsler para definir as estruturas de informação e seus α -sprays associados, bem como dualidade de métricas de Finsler e divergências afins.

3.1 Divergência: Definição.

Considere \mathcal{F} um conjunto de objetos que podem ser imagens $2D/3D$, distribuições de probabilidade, etc... Podemos calcular a diferença entre dois objetos de \mathcal{F} definindo uma função chamada de **divergência** no espaço $\mathcal{F} \times \mathcal{F}$ dada por:

$$\mathcal{D} : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}$$

$$(p, q) \rightarrow \mathcal{D}(p, q)$$

com as seguintes propriedades:

- $\mathcal{D}(p, q) \geq 0$
- $\mathcal{D}(p, q) = 0 \Leftrightarrow p = q$.

O número $\mathcal{D}(p, q)$ mede a divergência de p para q . O par $(\mathcal{F}, \mathcal{D})$ é chamado de **espaço de divergência**. Para o espaço de divergência $(\mathcal{F}, \mathcal{D})$, \mathcal{F} nem sempre tem dimensão finita. Na prática, consideramos uma família de objetos em \mathcal{F} parametrizados em um domínio de \mathbb{R}^n . Essa família é chamada de **modelo** em $(\mathcal{F}, \mathcal{D})$. Sendo mais preciso, um modelo de um espaço de divergência $(\mathcal{F}, \mathcal{D})$ é uma variedade M n -dimensional de classe C^∞ mergulhada em um subconjunto de \mathcal{F} com divergência induzida $D = \mathcal{D}|_M$. Esse modelo (M, D) também é um espaço de divergência.

Definição 3.1 Uma divergência \mathcal{D} é chamada **reversível**, se $\mathcal{D}(p, q) = \mathcal{D}(q, p)$.

Observação 5 \mathcal{D} nem sempre é reversível, ou seja, em geral, $\mathcal{D}(p, q) \neq \mathcal{D}(q, p)$.

Exemplo 3.2

Seja (\mathcal{M}, d) um espaço métrico. Então logicamente $\mathcal{D} := \frac{1}{2}d^2$ é uma divergência, pois $d : \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$ é uma função que satisfaz para $\forall x, y, z \in \mathcal{M}$:

- $d(x, x) = 0$,
- se $x \neq y \Rightarrow d(x, y) \geq 0$,
- $d(x, y) = d(y, x)$,
- $d(x, z) \leq d(x, y) + d(y, z)$.

Logo se $\mathcal{D} := \frac{1}{2}d^2$ então $\mathcal{D}(p, q) \geq 0$, pois $d(p, q) \geq 0$ e $\mathcal{D}(p, q) = 0 \Leftrightarrow p = q$ pois as duas primeiras propriedades de d nos dão $d(p, q) = 0 \Leftrightarrow p = q$. Esse é um exemplo de divergência reversível.

Exemplo 3.3

Seja $\Omega \subset \mathbb{R}^n$ um conjunto aberto. Seja $\Psi : \Omega \times \Omega \rightarrow \mathbb{R}$ dada por $\Psi = \Psi(x, y)$ com $\Psi(x, y) > \Psi(x, z) \forall y \neq z$. Seja $\Phi : \mathbb{R} \rightarrow \mathbb{R}$ uma função dada por $\Phi = \Phi(h)$, onde $\Phi(h) > 0$ para $h > 0$ e $\Phi(0) = 0$. Então a função $D(x, y) = \Phi(\Psi(x, y) - \Psi(x, x))$ é uma divergência em Ω , pois se $x \neq y$, então $\Psi(x, y) - \Psi(x, x) > 0$, logo $D(x, y) = \Phi(\Psi(x, y) - \Psi(x, x)) > 0$ e $D(x, y) = 0 \Leftrightarrow \Phi(\Psi(x, y) - \Psi(x, x)) = 0 \Leftrightarrow \Psi(x, x) = \Psi(x, y) \Leftrightarrow y = x$.

3.2 Divergências Regulares

Seja $d = d(p, q)$ a função distância de uma métrica de Finsler L em M . Seja

$$D(p, q) := \frac{1}{2}d(p, q)^2, \quad p, q \in M.$$

D é a divergência da métrica de Finsler em M . Em geral, a divergência D não é C^∞ ao longo da diagonal $\Delta = \{(p, p) \in M \times M\}$ a não ser que L seja Riemanniana. No entanto, temos o seguinte Lema, que, nessa dissertação será assumido como resultado.

Lema 3.4 *Se D é a divergência de uma métrica de Finsler L em uma variedade M , então para todo ponto p de M , existe um sistema de coordenadas locais (U, Φ) em M tal que*

$$2D(\Phi^{-1}(x), \Phi^{-1}(x+y)) = L(x, y) + \frac{1}{2}L_{x^k}(x, y)y^k + o(|y|^3). \quad (3-1)$$

Definição 3.5 *Seja M uma variedade. Uma função divergência D em M é chamada regular se em qualquer sistema de coordenadas locais (U, Φ) e todo ponto em M , restringindo a um pequeno domínio se necessário,*

$$2D(\Phi^{-1}(x), \Phi^{-1}(x+y)) = L(x, y) + P(x, y) + o(|y|^3), \quad (3-2)$$

onde $L = L(x, y)$ é uma métrica de Finsler em U e $P = P(x, y) \in C^\infty$ é uma função em $TU - \{0\}$ com

$$P(x, \lambda y) = \lambda^3 P(x, y), \lambda > 0.$$

A métrica de Finsler L na equação (3-2) forma uma métrica de Finsler global em M enquanto P não é uma função escalar global no fibrado TM . Contudo, podemos utilizar a função P para definir uma H -função em M .

Lema 3.6 *Seja D uma divergência regular em M . Seja L e P as funções locais definidas na equação (3-2) em um sistema de coordenadas locais (U, Φ) . Então*

$$H := P(x, y) - \frac{1}{2} L_{x^k}(x, y) y^k \quad (3-3)$$

é uma H -função bem definida em M .

Prova. Seja $\bar{L}(\bar{x}, \bar{y})$ e $\bar{P}(\bar{x}, \bar{y})$ as funções locais definidas em (3-2) em um sistema de coordenadas $(\bar{U}, \bar{\Phi})$. Seja $\bar{x} = \bar{\Phi} \circ \Phi^{-1}$. Então, expandindo-se em série de Taylor a expressão de \bar{x} , temos:

$$\bar{x}(x+y) = \bar{x} + \bar{y} + \frac{1}{2} \frac{\partial^2 \bar{x}(x)}{\partial x^i \partial x^j} y^i y^j + o(|y|^2), \quad (3-4)$$

onde

$$\bar{y} = \frac{\partial \bar{x}}{\partial x^i} y^i. \quad (3-5)$$

Reescrevamos a expressão (3-2) como

$$\begin{aligned} 2D(\Phi^{-1}(x), \Phi^{-1}(x+y)) &= 2D(\bar{\Phi}^{-1} \circ \bar{\Phi} \circ \Phi^{-1}(x), \bar{\Phi}^{-1} \circ \bar{\Phi} \circ \Phi^{-1}(x+y)) \\ &= 2D(\bar{\Phi}^{-1} \circ \bar{x}(x), \bar{\Phi}^{-1} \circ \bar{x}(x+y)) \\ &= 2D(\bar{\Phi}^{-1}(\bar{x}), \bar{\Phi}^{-1}(\bar{x} + \bar{z})), \end{aligned}$$

onde

$$\bar{z} = \bar{y} + \frac{1}{2} \frac{\partial^2 \bar{x}(x)}{\partial x^i \partial x^j} y^i y^j + o(|y|^2). \quad (3-6)$$

Mas D é uma divergência regular, então, por definição,

$$2D(\bar{\Phi}^{-1}(\bar{x}), \bar{\Phi}^{-1}(\bar{x} + \bar{z})) = \bar{L}(\bar{x}, \bar{z}) + \bar{P}(\bar{x}, \bar{z}) + o(|\bar{y}|^3), \quad (3-7)$$

onde

$$\begin{aligned} \bar{L}(\bar{x}, \bar{z}) &= \bar{L}(\bar{x}, \bar{y} + \frac{1}{2} \frac{\partial^2 \bar{x}(x)}{\partial x^i \partial x^j} y^i y^j + o(|y|^2)) \\ &= \bar{L}(\bar{x}, \bar{y}) + \frac{1}{2} \bar{L}_{y^k}(\bar{x}, \bar{y}) \frac{\partial^2 \bar{x}(x)}{\partial x^i \partial x^j} y^i y^j + o(|y|^3), \end{aligned}$$

e, analogamente,

$$\bar{P}(\bar{x}, \bar{z}) = \bar{P}(\bar{x}, \bar{y}) + o(|y|^3).$$

Então teremos, no novo sistema de coordenadas, a seguinte expansão:

$$\begin{aligned} 2D(\Phi^{-1}(x), \Phi^{-1}(x+y)) &= 2D(\bar{\Phi}^{-1} \circ \bar{x}(x), \bar{\Phi}^{-1} \circ \bar{x}(x+y)) = \\ &= \bar{L}(\bar{x}, \bar{y}) + \frac{1}{2} \bar{L}_{\bar{y}^k}(\bar{x}, \bar{y}) \frac{\partial^2 \bar{x}(x)}{\partial x^i \partial x^j} y^i y^j + \bar{P}(\bar{x}, \bar{y}) + o(|\bar{y}|^3). \end{aligned} \quad (3-8)$$

Comparando (3-2) e (3-8), teremos

$$L(x, y) = \bar{L}(\bar{x}, \bar{y}), \quad (3-9)$$

e

$$P(x, y) = \bar{P}(\bar{x}, \bar{y}) + \frac{1}{2} \bar{L}_{\bar{y}^k}(\bar{x}, \bar{y}) \frac{\partial^2 \bar{x}(x)}{\partial x^i \partial x^j} y^i y^j. \quad (3-10)$$

Diferenciando (3-9) em relação a x^k , aplicando y^k e multiplicando toda a equação por $\frac{1}{2}$, teremos:

$$\frac{1}{2} L_{x^k}(x, y) y^k = \frac{1}{2} \bar{L}_{\bar{x}^k}(\bar{x}, \bar{y}) \frac{\partial \bar{x}^k}{\partial x^i} y^i + \frac{1}{2} \bar{L}_{\bar{y}^k}(\bar{x}, \bar{y}) \frac{\partial \bar{x}^k}{\partial x^i} y^i.$$

Utilizando a equação (3-5), teremos

$$\frac{1}{2} L_{x^k}(x, y) y^k = \frac{1}{2} \bar{L}_{\bar{x}^k}(\bar{x}, \bar{y}) \bar{y}^k + \frac{1}{2} \bar{L}_{\bar{y}^k}(\bar{x}, \bar{y}) \frac{\partial^2 \bar{x}^k}{\partial x^i \partial x^j} y^i y^j. \quad (3-11)$$

Subtraindo (3-11) de (3-10), obteremos

$$P(x, y) - \frac{1}{2} L_{x^k}(x, y) y^k = \bar{P}(\bar{x}, \bar{y}) - \frac{1}{2} \bar{L}_{\bar{x}^k}(\bar{x}, \bar{y}) \bar{y}^k.$$

Concluimos então que (3-3) é uma H -função bem definida em M . \square

Logo, para uma divergência regular D , temos a seguinte expansão local:

$$2D(\Phi^{-1}(x), \Phi^{-1}(x+y)) = L(x, y) + \frac{1}{2} L_{x^k}(x, y) y^k + H(x, y) + o(|y|^3). \quad (3-12)$$

Proposição 3.7 *Se D é uma divergência de uma métrica de Finsler em uma variedade M , então ela é regular com $H = 0$.*

Prova. Segue diretamente do Lema 3.3, da definição 3.4 e do Lema 3.5. \square

Exemplo 3.8 Sejam Ω um subconjunto aberto em um espaço de Minkowski $(\mathbb{R}^n, \|\cdot\|)$, $\Psi(y) = a_{ijk}y^jy^k$, e

$$D(x, z) = \frac{1}{2} \|z - x\|^2 + \frac{1}{2} \Psi(z - x), x, z \in \Omega.$$

Tomemos o sistema de coordenadas natural $\phi(x) = x$. Então temos $2D(x, x+y) = \|y\|^2 - \Psi(y)$. D é uma divergência regular com $L(x, y) = \|y\|^2$ e $H(x, y) = \Psi(y)$.

3.3 Estruturas de Informação

Por definição, toda divergência regular D em uma variedade M induz uma métrica de Finsler L e uma H -função. Essas expressões podem ser obtidas por

$$L(x, y) = \lim_{\epsilon \rightarrow 0^+} \frac{2D(c(0), c(\epsilon))}{\epsilon^2}, \quad (3-13)$$

$$H(x, y) = \lim_{\epsilon \rightarrow 0^+} \frac{2D(\sigma(0), \sigma(\epsilon)) - L(x, y)\epsilon^2}{\epsilon^3}, \quad (3-14)$$

onde $c(t)$ é uma curva arbitrária de classe C^1 em M com $c(0) = x$ e $\dot{c}(0) = y$, e $\sigma = \sigma(t)$ é uma geodésica com $\sigma(0) = x$ e $\dot{\sigma}(0) = y$.

Definição 3.9 Uma *estrutura de informação* em uma variedade M é o par $\{L, H\}$, onde $L = L(x, y)$ é uma métrica de Finsler em M e $H = H(x, y)$ é uma H -função.

Toda divergência regular induz uma estrutura de informação, através do desenvolvimento em Séries de Taylor da mesma.

3.4 α -sprays de uma Estrutura de Informação

Seja (L, H) uma estrutura de informação em uma variedade M . Seja

$$\mathcal{G} = y^i \frac{\partial}{\partial x^i} - 2\mathcal{G}^i \frac{\partial}{\partial y^i} \quad (3-15)$$

um spray de L , onde

$$\mathcal{G}^i(x, \lambda y) = \lambda^2 \mathcal{G}^i(x, y),$$

para $\lambda > 0$ com

$$\mathcal{G}^i(x, y) = \frac{1}{4} g^{il}(x, y) \{L_{x^l y^i}(x, y) y^k - L_{x^i}(x, y)\}.$$

Utilizando H , vamos definir uma família de sprays

$$G_\alpha = y^i \frac{\partial}{\partial x^i} - 2G_\alpha^i \frac{\partial}{\partial y^i}, \quad (3-16)$$

onde

$$G_\alpha^i(x, y) = \mathcal{G}^i(x, y) + \frac{\alpha}{2} g^{ij}(x, y) H_{y^j}(x, y), \quad (3-17)$$

ou seja, vamos perturbar o spray de L . G_α é chamado de α -spray de (L, H) .

Lema 3.10 *Seja D uma divergência regular em uma variedade M , (L, H) a estrutura de informação induzida por D e G_α o α -spray de (L, H) . Seja $\sigma = \sigma(t)$ uma geodésica. Então para toda geodésica σ do spray G_α ,*

$$2D(\sigma(t_0), \sigma(t_0 + \varepsilon)) = L(x, y)\varepsilon^2 + (1 - 3\alpha)H(x, y)\varepsilon^3 + o(\varepsilon^3), \quad (3-18)$$

e

$$d(\sigma(t_0), \sigma(t_0 + \varepsilon))^2 = L(x, y)\varepsilon^2 - 3\alpha H(x, y)\varepsilon^3 + o(\varepsilon^3), \quad (3-19)$$

onde $\varepsilon > 0$, $x = \sigma(t_0)$, $y = \dot{\sigma}(t_0)$.

Prova. Considere $\phi = (x^i)$ um sistema de coordenadas locais em M . Seja $x(t) = \phi(\sigma(t))$ e $\Delta x = x(t_0 + \varepsilon) - x(t_0)$ usando uma expansão em série de Taylor, temos que

$$\begin{aligned} \Delta x^i &= x(t_0 + \varepsilon) - x(t_0) \\ &= \dot{x}^i(t_0)\varepsilon + \frac{1}{2}\ddot{x}^i(t_0)\varepsilon^2 + o(\varepsilon^2). \end{aligned} \quad (3-20)$$

Agora, levando em consideração o fato de que σ é geodésica, obtemos

$$\ddot{x}^i(t) + 2G_\alpha^i(x(t), \dot{x}(t)) = 0$$

logo

$$\begin{aligned} \Delta x^i &= \dot{x}^i(t_0)\varepsilon + \frac{1}{2}(-2G_\alpha^i)\varepsilon^2 + o(\varepsilon^2) \\ &= \dot{x}^i(t_0)\varepsilon - G_\alpha^i\varepsilon^2 + o(\varepsilon^2) \\ &= y^i\varepsilon - G_\alpha^i\varepsilon^2 + o(\varepsilon^2). \end{aligned} \quad (3-21)$$

Utilizando então (3-21), expansão em série de Taylor e o fato de que $L(x, y)$ e $H(x, y)$ são funções homogêneas positivas de grau 2 e 3 respectivamente em y , chegamos às seguintes expansões:

$$L(x, \Delta x) = \varepsilon^2 L - \varepsilon^3 L_{y^k} G_\alpha^k + o(\varepsilon^3), \quad (3-22)$$

$$L_{x^k}(x, \Delta x) \Delta x^k = \varepsilon^3 L_{x^k} y^k + o(\varepsilon^3), \quad (3-23)$$

$$H(x, \Delta x) = H\varepsilon^3 + o(\varepsilon^3). \quad (3-24)$$

Outros fatos também devem ser observados, o primeiro é que L_{y^k} é homogênea positiva de grau 1, devido ao fato de L ser homogênea positiva de grau 2 em y , logo pelo Teorema de Euler para funções homogêneas,

$$L_{y^k} = y^s L_{y^k y^s}. \quad (3-25)$$

O segundo é que

$$\mathcal{G}^k(x, y) = \frac{1}{4} g^{kl} \{L_{x^s y^l} y^s - L_{x^s}\}. \quad (3-26)$$

Então, podemos observar, que utilizando esses dois fatos,

$$\begin{aligned} L_{y^k} \mathcal{G}^k &= y^s L_{y^k y^s} \frac{1}{4} g^{kl} \{L_{x^s y^l} y^s - L_{x^s}\} \\ &= \frac{1}{4} y^s L_{y^k y^s} g^{kl} \{L_{x^s y^l} y^s - L_{x^s}\}. \end{aligned} \quad (3-27)$$

Mas sabemos que por definição de Métrica Finsler,

$$g_{ks} = \frac{1}{2} L_{y^k y^s}. \quad (3-28)$$

Então,

$$L_{y^k} \mathcal{G}^k = \frac{1}{2} y^s g_{ks} g^{kl} \{L_{x^s y^l} y^s - L_{x^s}\}. \quad (3-29)$$

Mas sabemos que

$$g_{ks} g^{kl} = \delta_{sl}, \quad (3-30)$$

então fazendo $s = l$, temos

$$L_{y^k} \mathcal{G}^k = \frac{1}{2} y^l \{L_{x^s y^l} y^s - L_{x^s}\}. \quad (3-31)$$

Observemos também que L_{x^s} é uma função homogênea positiva de grau 2 em y logo pelo Teorema de Euler para funções homogêneas,

$$y^l L_{x^s y^l} = 2L_{x^s}. \quad (3-32)$$

Sendo assim, a equação (3-31) fica

$$L_{y^k} \mathcal{G}^k = \frac{1}{2} 2L_{x^s} y^s - \frac{1}{2} L_{x^l} y^l. \quad (3-33)$$

como os índices s e l variam de 1 a n , podemos fazer $l = s = k$, portanto

$$\begin{aligned} L_{y^k} \mathcal{G}^k &= L_{x^s} y^s - \frac{1}{2} L_{x^l} y^l \\ &= \frac{1}{2} L_{x^k} y^k. \end{aligned} \quad (3-34)$$

Assim obtemos:

$$L_{y^k} \mathcal{G}^k = \frac{1}{2} L_{x^k} y^k. \quad (3-35)$$

Esta equação é de extrema importância não só para demonstrar este Lema, mas também para demonstrar lemas posteriores. Agora, por hipótese, sabemos que a divergência D é regular, logo

$$\begin{aligned} 2D(\sigma(t_0), \sigma(t_0 + \varepsilon)) &= 2D(\phi^{-1}(x), \phi^{-1}(x + \Delta x)) \\ &= L(x, \Delta x) + \frac{1}{2} L_{x^k}(x, \Delta x) \Delta x^k + H(x, \Delta x). \end{aligned} \quad (3-36)$$

Utilizando (3-22), (3-23) e (3-24), temos

$$2D(\sigma(t_0), \sigma(t_0 + \varepsilon)) = \varepsilon^2 L - \varepsilon^3 L_{y^k} G_{\alpha}^k + \frac{1}{2} \varepsilon^3 L_{x^k} y^k + \varepsilon^3 H + o(\varepsilon^3). \quad (3-37)$$

Utilizando a equação (3-35), obteremos

$$\begin{aligned} 2D(\sigma(t_0), \sigma(t_0 + \varepsilon)) &= \varepsilon^2 L - \varepsilon^3 L_{y^k} G_{\alpha}^k + \varepsilon^3 L_{y^k} \mathcal{G}^k + \varepsilon^3 H + o(\varepsilon^3) \\ &= \varepsilon^2 L - \varepsilon^3 L_{y^k} \{G_{\alpha}^k - \mathcal{G}^k\} + \varepsilon^3 H + o(\varepsilon^3). \end{aligned} \quad (3-38)$$

Utilizando a equação (3-17) temos

$$\begin{aligned} 2D(\sigma(t_0), \sigma(t_0 + \varepsilon)) &= \varepsilon^2 L - \varepsilon^3 L_{y^k} \frac{\alpha}{2} g^{kj} H_{y^j} + \varepsilon^3 H + o(\varepsilon^3) \\ &= \varepsilon^2 L - \varepsilon^3 \frac{\alpha}{2} L_{y^k} g^{kj} H_{y^j} + \varepsilon^3 H + o(\varepsilon^3). \end{aligned} \quad (3-39)$$

Utilizando a equação (3-25), obtemos

$$2D(\sigma(t_0), \sigma(t_0 + \varepsilon)) = \varepsilon^2 L - \varepsilon^3 \frac{\alpha}{2} y^m L_{y^k y^m} g^{kj} H_{y^j} + \varepsilon^3 H + o(\varepsilon^3). \quad (3-40)$$

Usando a equação (3-28), e fato semelhante à equação (3-30) temos

$$\begin{aligned} 2D(\sigma(t_0), \sigma(t_0 + \varepsilon)) &= \varepsilon^2 L - \varepsilon^3 \alpha y^m g_{km} g^{kj} H_{y^j} + \varepsilon^3 H + o(\varepsilon^3) \\ &= \varepsilon^2 L - \varepsilon^3 \alpha y^m \delta_{mj} H_{y^j} + \varepsilon^3 H + o(\varepsilon^3) \\ &= \varepsilon^2 L - \varepsilon^3 \alpha y^j H_{y^j} + \varepsilon^3 H + o(\varepsilon^3). \end{aligned} \quad (3-41)$$

Vamos agora utilizar o Teorema de Euler pra funções homogêneas. Como H é uma função positivamente homogênea de grau 3, temos

$$y^j H_{y^j} = 3H, \quad (3-42)$$

logo,

$$2D(\sigma(t_0), \sigma(t_0 + \varepsilon)) = \varepsilon^2 L - \varepsilon^3 \alpha (3H) + \varepsilon^3 H + o(\varepsilon^3), \quad (3-43)$$

onde obtemos finalmente

$$2D(\sigma(t_0), \sigma(t_0 + \varepsilon)) = \varepsilon^2 L + (1 - 3\alpha)H\varepsilon^3 + o(\varepsilon^3). \quad (3-44)$$

O Lema 3.3, nos dá a expressão para o desenvolvimento em séries de Taylor, para a divergência de uma métrica de Finsler L , dada por

$$D(p, q) := \frac{1}{2} d(p, q)^2.$$

Nas condições das hipóteses desse Lema, temos:

$$\begin{aligned} d(\sigma(t_0), \sigma(t_0 + \varepsilon))^2 &= d(\Phi^{-1}(x), \Phi^{-1}(x + \Delta x))^2 \\ &= L(x, \Delta x) + \frac{1}{2} L_{x^k}(x, \Delta x) \Delta x^k + o(\Delta x^3). \end{aligned}$$

Utilizando-se dos desenvolvimentos em séries de Taylor (3-22) e (3-23), teremos

$$d(\sigma(t_0), \sigma(t_0 + \varepsilon))^2 = L\varepsilon^2 + L_{y^k} \varepsilon^3 (\mathcal{G}^k - G_\alpha^k) + o(\varepsilon^3).$$

Fazendo uma analogia à demonstração da equação (3-18), teremos que a equação acima

se reduz a:

$$\begin{aligned}
 d(\sigma(t_0), \sigma(t_0 + \varepsilon))^2 &= L\varepsilon^2 - L_{y^k} \varepsilon^3 \frac{\alpha}{2} g^{kj} H_{y^j} + o(\varepsilon^3) \\
 &= L\varepsilon^2 - \frac{\alpha}{2} \varepsilon^3 y^j L_{y^k y^j} g^{kj} H_{y^j} + o(\varepsilon^3) \\
 &= L\varepsilon^2 - \frac{\alpha}{2} \varepsilon^3 y^j 2g_{kl} g^{kj} H_{y^j} + o(\varepsilon^3) \\
 &= L\varepsilon^2 - \alpha \varepsilon^3 y^j \delta_{lj} H_{y^j} + o(\varepsilon^3) \\
 &= L\varepsilon^2 - \alpha \varepsilon^3 y^j H_{y^j} + o(\varepsilon^3) \\
 &= L\varepsilon^2 - 3\alpha H \varepsilon^3 + o(\varepsilon^3).
 \end{aligned}
 \tag{3-45}$$

Concluimos então a demonstração do Lema. \square

Observação 6 Da prova do Lema anterior, podemos tirar a seguinte conclusão: O α -spray associado para $\alpha = \frac{1}{3}$ é o "melhor" spray, pelo fato de que ao longo de toda geodésica dele, a divergência muda a taxa mínima, ou seja, no caso $\alpha = \frac{1}{3}$, temos a seguinte configuração:

$$2D(\sigma(t_0), \sigma(t_0 + \varepsilon)) = \varepsilon^2 L + o(\varepsilon^3).$$

Ao longo geodésicas de $G_{\frac{1}{3}}$ a variação de D é mínima em relação à métrica.

Definição 3.11 Uma estrutura de informação (L, H) em uma variedade M é chamada de α -flat para algum α se o α -spray G_α de (L, H) é flat. (L, H) é chamada de flat se é 1-flat.

Definição 3.12 Seja (L, H) uma estrutura de informação em uma variedade M . Seja $L^*(x, y) = L(x, -y)$ e $H^*(x, y) = H(x, -y)$. Então (L^*, H^*) também é uma estrutura de informação em M chamada estrutura de informação dual de (L, H) .

Lema 3.13 Seja (L, H) uma estrutura de informação em uma variedade M . Então :

- (i) (L, H) é α -flat se e somente se $(L, \alpha H)$ é 1-flat;
- (ii) (L, H) é α -flat se e somente se (L^*, H^*) é $(-\alpha)$ -flat.

Prova.

(i): Se (L, H) é α -flat temos equivalentemente que $G_\alpha^i = 0$. Mas esse fato acontece se, e somente se,

$$G^i(x, y) = -\frac{\alpha}{2} g^{ij}(x, y) H_{y^j}(x, y) = -\frac{1}{2} g^{ij}(x, y) (\alpha H)_{y^j}(x, y),$$

fato equivalente a $(L, \alpha H)$ ser 1-flat.

(ii) Observemos o fato de que $L^*(x, y) = L(x, -y)$ então $L_{y^j}^*(x, y) = -L_{y^j}(x, -y)$ e $L_{x^k}^*(x, y) = L_{x^k}(x, -y)$, sendo assim $L_{x^k y^j}^*(x, y) = -L_{x^k y^j}(x, -y)$. Logo

$$g_{ij}^*(x, y) = \frac{1}{2} L_{y^i y^j}^*(x, y) = \frac{1}{2} L_{y^i y^j}(x, -y) = g_{ij}(x, -y) \quad (3-46)$$

e

$$\mathcal{G}^{*i}(x, y) = \frac{1}{4} g^{ij}(x, -y) \{-L_{x^k y^j}(x, -y) - L_{x^k}(x, y)\} = \mathcal{G}^i(x, -y). \quad (3-47)$$

Temos também que

$$H^*(x, -y) = H(x, -y) \Rightarrow H_{y^j}^*(x, y) = -H_{y^j}(x, -y), \quad (3-48)$$

assim

$$\begin{aligned} G_{\alpha}^i(x, y) &= \mathcal{G}^i(x, -y) + \frac{\alpha}{2} g^{ij}(x, -y) H_{y^j}(x, -y) \\ &= \mathcal{G}^{*i}(x, y) - \frac{\alpha}{2} g^{*ij}(x, y) H_{y^j}^*(x, y) \\ &= G_{-\alpha}^{*i}(x, y). \end{aligned} \quad (3-49)$$

Em posse desses fatos, então (L, H) é α -flat se, e somente se, $G_{\alpha}^i(x, y) = 0 \forall (x, y) \in TM$. Então $G_{\alpha}^i(x, -y) = 0$ mas $G_{\alpha}^i(x, -y) = G_{-\alpha}^{*i}(x, y)$ por (2.39). Logo $G_{-\alpha}^{*i}(x, y) = 0$ se, e somente se, (L^*, H^*) é $-\alpha$ -flat. \square

Lema 3.14 *Seja (L, H) uma estrutura de informação em uma variedade M . Para algum $\alpha \neq 0$, (L, H) é α -flat se, e somente se, para todo ponto existe um sistema de coordenadas locais (x^i) tal que*

$$L_{x^k y^j} y^k = 2L_{x^j}, \quad (3-50)$$

$$\alpha H = -\frac{1}{6} L_{x^k} y^k. \quad (3-51)$$

Prova. Suponhamos que (L, H) seja α -flat. Assuma que exista um sistema de coordenadas (x^i, y^j) em que $G_{\alpha}^i(x, y) = 0$. Utilizando (3-34), (3-25) e o fato de que (L, H) é α -flat se, e somente se, $G_{\alpha}^k = 0$ o que é equivalente a

$$\mathcal{G}^k(x, y) = -\frac{\alpha}{2} g^{kj}(x, y) H_{y^j}(x, y), \quad (3-52)$$

$$\begin{aligned}\frac{1}{2}L_{x^k}y^k &= L_{y^k}\mathcal{G}^k \\ &= \left(y^m L_{y^k y^m}\right)\left(-\frac{\alpha}{2}g^{kj}H_{y^j}\right).\end{aligned}\tag{3-53}$$

Agora, utilizando (3-28), temos

$$\begin{aligned}L_{x^k}y^k &= -\alpha y^m L_{y^k y^m}g^{kj}H_{y^j} \\ &= -2\alpha y^m g_{km}g^{kj}H_{y^j}.\end{aligned}\tag{3-54}$$

Utilizando um fato análogo à (3-30) obtemos

$$\begin{aligned}L_{x^k}y^k &= -2\alpha y^m \delta_{mj}H_{y^j} \\ &= -2\alpha y^j H_{y^j}.\end{aligned}\tag{3-55}$$

Utilizando a equação (3-42) temos

$$L_{x^k}y^k = -2\alpha 3H,$$

logo

$$\alpha H = -\frac{1}{6}L_{x^k}y^k.\tag{3-56}$$

Alterando o nome de alguns índices em (3-52) temos por hipótese que

$$\begin{aligned}\mathcal{G}^i(x, y) &= -\frac{\alpha}{2}g^{il}(x, y)H_{y^l}(x, y) \\ &= -\frac{1}{2}g^{il}(\alpha H_{y^l}).\end{aligned}\tag{3-57}$$

Utilizando (3-56), obteremos

$$\begin{aligned}\mathcal{G}^i(x, y) &= -\frac{1}{2}g^{il}(\alpha H_{y^l}) \\ &= \frac{1}{2}g^{il}\left(\frac{1}{6}L_{x^k}y^k\right)_{y^l}.\end{aligned}\tag{3-58}$$

Diferenciando a expressão entre parênteses e simplificando o que for possível , temos

$$\begin{aligned}
\mathcal{G}^i(x,y) &= \frac{1}{12}g^{il} \left(L_{x^k y^l} y^k + L_{x^k} \frac{\partial y^k}{\partial y^l} \right) \\
&= \frac{1}{12}g^{il} \left(L_{x^k y^l} y^k + L_{x^k} \delta_{kl} \right) \\
&= \frac{1}{12}g^{il} \left(L_{x^k y^l} y^k + L_{x^l} \right).
\end{aligned} \tag{3-59}$$

Por outro lado,

$$\mathcal{G}^i(x,y) = \frac{1}{4}g^{il} \left(L_{x^k y^l} y^k - L_{x^l} \right), \tag{3-60}$$

então

$$\frac{1}{12}g^{il} \left(L_{x^k y^l} y^k + L_{x^l} \right) = \frac{1}{4}g^{il} \left(L_{x^k y^l} y^k - L_{x^l} \right). \tag{3-61}$$

Multiplicando ambos os termos por g_{is} e utilizando fato análogo a (3-30), temos que

$$\frac{1}{3}g_{is}g^{il} \left(L_{x^k y^l} y^k + L_{x^l} \right) = g_{is}g^{il} \left(L_{x^k y^l} y^k - L_{x^l} \right) \tag{3-62}$$

$$\frac{1}{3}\delta_{sl} \left(L_{x^k y^l} y^k + L_{x^l} \right) = \delta_{sl} \left(L_{x^k y^l} y^k - L_{x^l} \right) \tag{3-63}$$

$$\frac{1}{3} \left(L_{x^k y^l} y^k + L_{x^l} \right) = \left(L_{x^k y^l} y^k - L_{x^l} \right), \tag{3-64}$$

logo,

$$L_{x^k y^l} y^k + L_{x^l} = 3L_{x^k y^l} y^k - 3L_{x^l}. \tag{3-65}$$

Desse fato, obtemos:

$$L_{x^k y^l} y^k = 2L_{x^l}. \tag{3-66}$$

Por outro lado, se L satisfaz (3-50) então

$$\begin{aligned}
\mathcal{G}^i(x,y) &= \frac{1}{4}g^{il} \left(L_{x^k y^l} y^k - L_{x^l} \right) \\
&= \frac{1}{4}g^{il} (2L_{x^l} - L_{x^l}) \\
&= \frac{1}{4}g^{il} L_{x^l}.
\end{aligned} \tag{3-67}$$

Se L satisfaz (3-51) temos

$$\begin{aligned}
 (\alpha H)_{y^j} &= \left(\frac{1}{6} L_{x^k} y^k \right)_{y^j} \\
 &= \left(-\frac{1}{6} L_{x^k y^j} y^k - \frac{1}{6} L_{x^j} \right) \\
 &= \left(-\frac{1}{6} \cdot 2L_{x^j} - \frac{1}{6} L_{x^j} \right) \\
 &= -\frac{1}{2} L_{x^j}.
 \end{aligned} \tag{3-68}$$

Então

$$\begin{aligned}
 \frac{\alpha}{2} g^{ij} H_{y^j} &= \frac{1}{2} g^{ij} (\alpha H)_{y^j} \\
 &= \frac{g^{ij}}{2} \left(-\frac{1}{2} L_{x^j} \right) \\
 &= -\frac{1}{4} g^{ij} L_{x^j}.
 \end{aligned} \tag{3-69}$$

Logo

$$G_{\alpha}^i(x, y) = G^i(x, y) + \frac{\alpha}{2} g^{ij} H_{y^j} = \frac{1}{4} g^{ij} L_{x^j} - \frac{1}{4} g^{ij} L_{x^j} = 0. \tag{3-70}$$

Portanto (L, H) é α -flat. \square

3.5 Métricas Finsler Dualmente Flat

Definição 3.15 Uma métrica Finsler L é chamada **Localmente Dualmente Flat** se para todo ponto existe um sistema de coordenadas (x^i) tal que $L = L(x, y)$ satisfaz (3-50). Esse sistema de coordenadas é chamado de **Sistema Local Adaptado**. L é chamada de **Globalmente Dualmente Flat** se existe uma H -função tal que a estrutura de informação (L, H) é 1-flat, isto é, que para todo ponto existe um sistema de coordenadas locais tal que L satisfaz (3-51).

Se L é uma métrica de Finsler localmente dualmente flat em uma variedade M , então em todo ponto existe um sistema de coordenadas (x^i) no qual o spray com coeficientes $(G)^i$ é 1-flat, ou seja

$$G^i(x, y) + \frac{1}{2} g^{ij}(x, y) H_{y^j}(x, y) = 0, \tag{3-71}$$

onde

$$H = -\frac{1}{6}L_{x^k}y^k. \quad (3-72)$$

A Proposição a seguir considera as métricas Riemannianas Localmente Dualmente Flat.

Proposição 3.16 *Uma métrica Riemanniana $L = g_{ij}(x)y^i y^j$ em uma variedade M é Localmente Dualmente Flat se e somente se ela pode ser localmente expressa como*

$$g_{ij}(x) = \frac{\partial^2 \Psi}{\partial x^i \partial x^j}(x), \quad (3-73)$$

onde $\Psi = \Psi(x)$ é uma função escalar local em M .

Prova. A volta da Proposição é simples: se $g_{ij}(x) = \frac{\partial^2 \Psi}{\partial x^i \partial x^j}(x)$, então considere L a métrica Riemanniana dada por

$$L = g_{ij}(x)y^i y^j = \frac{\partial^2 \Psi}{\partial x^i \partial x^j}(x)y^i y^j. \quad (3-74)$$

Derivando-se (3-74), em relação a x^l e a x^k respectivamente, temos as seguintes expressões:

$$L_{x^l} = \frac{\partial^3 \Psi}{\partial x^l \partial x^i \partial x^j}(x)y^i y^j \quad (3-75)$$

e

$$L_{x^k} = \frac{\partial^3 \Psi}{\partial x^k \partial x^i \partial x^j}(x)y^i y^j. \quad (3-76)$$

Temos também que, derivando (3-76) em relação a y^l e somando em y^k , temos

$$L_{x^k y^l} y^k = \frac{\partial^3 \Psi(x)}{\partial x^k \partial x^i \partial x^j} (y^i \delta_{lj} + y^j \delta_{li}) y^k. \quad (3-77)$$

Efetuada as simplificações necessárias, temos

$$L_{x^k y^l} y^k = \frac{\partial^3 \Psi(x)}{\partial x^k \partial x^i \partial x^l} y^i y^k + \frac{\partial^3 \Psi(x)}{\partial x^k \partial x^l \partial x^i} y^i y^k. \quad (3-78)$$

Como os índices i e j têm a mesma variação e trata-se de uma soma, faça $j = i$ no segundo termo da soma anterior, assim, teremos que

$$L_{x^k y^l} y^k = 2 \frac{\partial^3 \Psi(x)}{\partial x^k \partial x^i \partial x^l} y^i y^k = 2L_{x^l}. \quad (3-79)$$

Assim L é localmente dualmente flat. Por outro lado, se $L = g_{ij}(x)y^i y^j$ ou seja, Riemanniana e é localmente dualmente flat, ou seja, obedece à equação (3-50), temos primeira-

mente derivando L em relação a x^l que

$$L_{x^l} = \frac{\partial g_{ij}}{\partial x^l} y^i y^j, \quad (3-80)$$

e em relação a x^k que

$$L_{x^k} = \frac{\partial g_{ij}}{\partial x^k} y^i y^j. \quad (3-81)$$

Derivando (3-81) em relação a y^l e somando em y^k temos que

$$L_{x^k y^l} y^k = \frac{\partial g_{lj}}{\partial x^k} y^j y^k + \frac{\partial g_{il}}{\partial x^k} y^i y^k. \quad (3-82)$$

Como temos uma soma em (3-82) podemos fazer uma troca de índices de modo que os 3 termos da equação fiquem em $y^i y^k$, pois todos os índices possuem a mesma variação. Sendo assim, chegamos a

$$\frac{\partial g_{li}}{\partial x^k}(x) + \frac{\partial g_{kl}}{\partial x^i}(x) = 2 \frac{\partial g_{ik}}{\partial x^l}(x). \quad (3-83)$$

Fazendo a permutação dos índices i e l em (3-83), temos

$$\frac{\partial g_{li}}{\partial x^k}(x) + \frac{\partial g_{ik}}{\partial x^l}(x) = 2 \frac{\partial g_{kl}}{\partial x^i}(x). \quad (3-84)$$

De onde obtemos

$$\frac{\partial g_{ik}}{\partial x^l}(x) = \frac{\partial g_{kl}}{\partial x^i}(x) = \frac{\partial g_{li}}{\partial x^k}(x). \quad (3-85)$$

Seja

$$\Psi(x) = \int \int g_{ii}(x) dx^i dx^i \quad (3-86)$$

então, temos que, derivando-se parcialmente Ψ em relação a x^k que

$$\frac{\partial \Psi}{\partial x^k}(x) = \int \left[\int \frac{\partial g_{il}}{\partial x^k} dx^l \right] dx^i. \quad (3-87)$$

Utilizando a equação (3-85) temos

$$\frac{\partial \Psi}{\partial x^k}(x) = \int g_{ik} dx^i. \quad (3-88)$$

Derivando toda a expressão em relação a x^l temos

$$\frac{\partial^2 \Psi}{\partial x^k \partial x^l}(x) = \int \frac{\partial g_{ik}}{\partial x^l} dx^i = g_{kl}. \quad (3-89)$$

□

Proposição 3.17 *Seja $\Omega \subset \mathbb{R}^n$ um domínio extritamente convexo definido por uma norma de Minkowski $\phi(y)$ em \mathbb{R}^n ,*

$$\Omega := \{y \in \mathbb{R}^n : \phi(y) < 1\}.$$

Defina $\theta(x, y) > 0$, $y \neq 0$ por

$$\theta(x, y) = \phi(y + \theta(x, y)x), y \in T_x\Omega = \mathbb{R}^n. \quad (3-90)$$

Seja $L(x, y) = (\theta(x, y))^2$. Então L é localmente dualmente flat.

Prova. Seja $w = y + \theta(x, y)x$. Diferenciando a equação (3-90) em relação a x^k e y^k respectivamente, temos

$$\theta_{x^k} = \phi_{w^k}(w)\theta(x, y) + \phi_{w^j}x^j\theta_{x^k}, \quad (3-91)$$

ou seja

$$(1 - \phi_{w^j}(w)x^j)\theta_{x^k} = \phi_{w^k}(w)\theta(x, y) \quad (3-92)$$

e

$$\theta_{y^k} = \phi_{w^k}(w) + \phi_{w^j}x^j\theta_{y^k}, \quad (3-93)$$

ou seja

$$(1 - \phi_{w^j}(w)x^j)\theta_{y^k} = \phi_{w^k}(w). \quad (3-94)$$

Comparando (3-92) e (3-94) chegamos à conclusão de que

$$\theta_{x^k}(x, y) = \theta(x, y)\theta_{y^k}(x, y). \quad (3-95)$$

Um fato que deve ser observado é que θ é definido por uma métrica de Minkowski logo é positivamente homogêneo de grau 1. Temos então que

$$\begin{aligned} L_{x^k} &= 2\theta\theta_{x^k} \\ &= 2\theta\theta_{y^k} \\ &= 2\theta^2\theta_{y^k}. \end{aligned} \quad (3-96)$$

Derivando (3-96) em relação a y^j e aplicando y^k temos

$$L_{x^k y^j} y^k = 4\theta_{y^j}\theta_{y^k}y^k + 2\theta^2\theta_{y^k y^j}. \quad (3-97)$$

Como θ é positivamente homogêneo de grau 1, então pelo Teorema de Euler para funções homogêneas,

$$\theta_{y^i} y^i = \theta \quad (3-98)$$

e

$$\theta_{y^i y^j} y^i y^j = 0, \forall i, j. \quad (3-99)$$

Utilizando esses dois fatos, concluímos que

$$L_{x^i y^j} y^i y^j = 4\theta^2 \theta_{y^j}. \quad (3-100)$$

Como

$$2L_{x^i} = 4\theta^2 \theta_{y^i}, \quad (3-101)$$

temos que (3-50) é satisfeita. Logo L é localmente dualmente flat. \square

3.6 Divergências afins e Estruturas de Informação afins

Em geral uma divergência regular D em uma variedade M não é uma função C^∞ ao longo da diagonal $M \times M$. Nessa Seção iremos estudar os casos particulares em que D é uma divergência C^∞ ao longo da diagonal.

Definição 3.18 Uma divergência regular D em uma variedade M é chamada **afim** se D é uma função C^∞ em uma vizinhança da diagonal $M \times M$.

Lema 3.19 Seja D uma divergência regular afim em uma variedade M . Então a informação induzida (L, H) satisfaz

$$(i) \quad L = g_{ij}(x) y^i y^j \text{ é Riemanniana.}$$

$$(ii) \quad H = H_{ijk}(x) y^i y^j y^k.$$

Prova.

Façamos a seguinte identificação: $D(x, z) := D(\phi^{-1}(x), \phi^{-1}(z))$ onde $\phi : M \rightarrow \mathbb{R}^n$ é um sistema de coordenadas. Vamos assumir que $D(x, z)$ é C^∞ em $(x, z) \in M \times M$. Lembrando-se que a divergência D é regular, temos então a seguinte expansão em séries de Taylor:

$$2D(x, x+y) = g_{ij}(x) y^i y^j + \frac{1}{3} h_{ijk}(x) y^i y^j y^k + o(|y|^3), \quad (3-102)$$

onde

$$g_{ij}(x) = \left[\frac{\partial^2 D(x, z)}{\partial z^i \partial z^j} \right]_{|z=x} \quad (3-103)$$

e

$$h_{ijk}(x) = \left[\frac{\partial^3 D(x, z)}{\partial z^i \partial z^j \partial z^k} \right]_{|_{\{x=z\}}} \quad (3-104)$$

Defina agora

$$H_{ijk}(x) = \frac{1}{3} h_{ijk}(x) - \frac{1}{6} \left[\frac{\partial g_{ij}}{\partial x^k}(x) + \frac{\partial g_{ik}}{\partial x^j}(x) + \frac{\partial g_{jk}}{\partial x^i}(x) \right] \quad (3-105)$$

Então podemos escrever (3-102) como

$$2D(x, x+y) = g_{ij}(x)y^i y^j + H_{ijk}(x)y^i y^j y^k + \frac{1}{6} \left[\frac{\partial g_{ij}}{\partial x^k}(x) + \frac{\partial g_{ik}}{\partial x^j}(x) + \frac{\partial g_{jk}}{\partial x^i}(x) \right] y^i y^j y^k + o(|y|^3) \quad (3-106)$$

Como estamos tendo agora uma soma em i, j e k , e esses índices tem a mesma variação, temos

$$\begin{aligned} \frac{1}{6} \left[\frac{\partial g_{ij}}{\partial x^k}(x) + \frac{\partial g_{ik}}{\partial x^j}(x) + \frac{\partial g_{jk}}{\partial x^i}(x) \right] y^i y^j y^k &= 3 \cdot \frac{1}{6} \frac{\partial g_{ij}}{\partial x^k} y^i y^j y^k \\ &= \frac{1}{2} \frac{\partial g_{ij}}{\partial x^k} y^i y^j y^k. \end{aligned} \quad (3-107)$$

Deste modo,

$$2D(x, x+y) = g_{ij}(x)y^i y^j + H_{ijk}(x)y^i y^j y^k + \frac{1}{2} \frac{\partial g_{ij}}{\partial x^k}(x)y^i y^j y^k + o(|y|^3) \quad (3-108)$$

Logo $L = g_{ij}(x)y^i y^j$, ou seja g_{ij} depende de $x \in M$, assim L é Riemanniana e $H = H_{ijk}(x)y^i y^j y^k$ é polinomial. \square

logo, podemos formular a seguinte definição:

Definição 3.20 Uma estrutura de informação (L, H) em uma variedade M é chamada de afim se L é Riemanniana e $H = H_{ijk}y^i y^j y^k$ é um polinômio homogêneo.Se (L, H) é uma estrutura de informação afim, então $(L^*, H^*) = (L, -H)$, pois se

$$L(x, y) = g_{ij}(x)y^i y^j$$

, então

$$L(x, -y) = g_{ij}(x)(-y)^i (-y)^j = g_{ij}(x)y^i y^j = L(x, y)$$

e, se

$$H(x, y) = H_{ijk}(x)y^i y^j y^k,$$

obtemos

$$H(x, -y) = H_{ij}(x)(-y)^i (-y)^j (-y)^k = -H_{ijk}(x)y^i y^j y^k = -H(x, y).$$

Lema 3.21 Para uma divergência afim em uma variedade M e seu dual D^* , a estrutura de informação (L, H) de D é dual à estrutura de informação (L^*, H^*) de D^* .

Prova. Reescrevamos a equação (3-108) como

$$\begin{aligned} 2D^*(x, x+y) &= g_{ij}^*(x)y^i y^j + H_{ijk}^*(x)y^i y^j y^k + \frac{1}{2} \frac{\partial g_{ij}^*}{\partial x^k}(x)y^i y^j y^k + o(|y|^3) \\ &= L^*(x, y) + H^*(x, y) + \frac{1}{2} \frac{\partial g_{ij}^*}{\partial x^k}(x)y^i y^j y^k + o(|y|^3) \\ &= L(x, y) - H(x, y) + \frac{1}{2} \frac{\partial g_{ij}^*}{\partial x^k}(x)y^i y^j y^k + o(|y|^3). \end{aligned}$$

Logo a estrutura de informação induzida por D^* é $(L, -H)$. \square

3.7 Estruturas de Informação afins do tipo α -Flat

Nessa seção estaremos particularmente interessados na estruturas de informação α -flat. Se uma estrutura de informação é α -flat, então o α -spray associado à esta estrutura é *flat*. Nessa seção daremos ênfase às estruturas de informação afins e α -flat, com o objetivo de provar que uma estrutura de informação afim (L, H) é α -flat se e somente se a estrutura de informação dual (L^*, H^*) é α -flat.

Lema 3.22 Seja (L, H) uma estrutura de informação afim em uma variedade M e $\alpha \neq 0$. Então (L, H) é α -flat se, e somente se, existe um sistema de coordenadas locais (x^i) e uma função local $\Psi = \Psi(x)$ tal que

$$L(x, y) = \frac{\partial^2 \Psi}{\partial x^i \partial x^j}(x) y^i y^j, \quad (3-109)$$

$$H(x, y) = -\frac{1}{6\alpha} \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k}(x) y^i y^j y^k. \quad (3-110)$$

Prova.

Assuma que (L, H) é uma estrutura de informação α -flat. Pelo Lema 3.4, temos que existe um sistema de coordenadas locais (x^i) tal que

$$L_{x^k y^j} y^k = 2L_{x^j}.$$

Utilizando a Proposição 3.16 concluímos que

$$g_{ij}(x) = \frac{\partial^2 \psi}{\partial x^i \partial x^j}(x),$$

onde $\psi = \psi(x)$ é uma função escalar local em M . Logo (3-109) é satisfeita.

Utilizando a equação (3-51), ou seja

$$\alpha H = -\frac{1}{6} L_{x^k} y^k,$$

temos que primeiramente, como $L = g_{ij}(x) y^i y^j$, então

$$\begin{aligned} \alpha H &= -\frac{1}{6} \frac{\partial}{\partial x^k} \left(\frac{\partial^2 \psi}{\partial x^i \partial x^j} \right) (x) y^i y^j y^k \\ &= -\frac{1}{6} \frac{\partial^3 \psi}{\partial x^i \partial x^j \partial x^k} (x) y^i y^j y^k. \end{aligned} \quad (3-111)$$

Assim temos:

$$H(x, y) = -\frac{1}{6\alpha} \frac{\partial^3 \psi}{\partial x^i \partial x^j \partial x^k} (x) y^i y^j y^k. \quad (3-112)$$

Inversamente, suponha agora que $L = g_{ij}(x) y^i y^j$ e $H = H_{ijk}(x) y^i y^j y^k$ satisfaçam as equações (3-109) e (3-110) respectivamente. Então L satisfaz (3-50), basta olhar a recíproca da Proposição 3.16. Se H satisfaz (3-121), então

$$\begin{aligned} -6\alpha H(x, y) &= \frac{\partial^3 \psi}{\partial x^i \partial x^j \partial x^k} (x) y^i y^j y^k \\ &= \frac{\partial}{\partial x^k} \left(\frac{\partial^2 \psi(x)}{\partial x^i \partial x^j} y^i y^j \right) y^k \\ &= \frac{\partial}{\partial x^k} (g_{ij}(x) y^i y^j) y^k \\ &= L_{x^k} y^k. \end{aligned} \quad (3-113)$$

assim H satisfaz (3-51). Logo, pelo Lema 3.4 temos que (L, H) é α -flat. \square

Lema 3.23 *Seja (L, H) uma estrutura de informação afim em uma variedade M e $\alpha \neq 0$. Assuma que em um sistema de coordenadas locais (x^i) , (L, H) é dado por (3-109) e*

(3-110) respectivamente. Seja

$$x_i^* = \frac{\partial \Psi}{\partial x^i}(x)$$

e

$$\Psi^*(x^*) := -\Psi(x) + \sum_{i=1}^n x_i^* x^i. \quad (3-114)$$

Então no novo sistema de coordenadas (x^{*i}) , a estrutura de informação dual $(L^*, H^*) = (L, -H)$ é dada por

$$L^*(x^*, y^*) = \frac{\partial^2 \Psi^*(x^*)}{\partial x_i^* \partial x_j^*} y_i^* y_j^* \quad (3-115)$$

e

$$H^*(x^*, y^*) = -\frac{1}{6\alpha} \frac{\partial^3 \Psi^*(x^*)}{\partial x_i^* \partial x_j^* \partial x_k^*} y_i^* y_j^* y_k^*. \quad (3-116)$$

Logo (L^*, H^*) é α -flat.

Prova.

Primeiramente, sabemos que os coeficientes do spray de uma métrica de Finsler L é dada por

$$G^i(x, y) = \frac{1}{4} g^{ik} \{L_{x^l y^k y^l} - L_{x^k}\} \quad (3-117)$$

Por hipótese, a estrutura de informação é afim, então

$$L(x, y) = \frac{\partial^2 \Psi}{\partial x^i \partial x^j}(x) y^i y^j$$

e

$$L_{x^l} = \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^l}(x) y^i y^j.$$

Assim

$$L_{x^l y^k y^l} = \frac{\partial^3 \Psi}{\partial x^i \partial x^l \partial x^k}(x) y^i y^l + \frac{\partial^3 \Psi}{\partial x^i \partial x^k \partial x^l}(x) y^i y^l. \quad (3-118)$$

Então, fazendo a troca de l por k na expressão de L_{x^l} , temos

$$\begin{aligned} L_{x^l y^k y^l} - L_{x^k} &= \frac{\partial^3 \Psi}{\partial x^i \partial x^l \partial x^k}(x) y^i y^l + \frac{\partial^3 \Psi}{\partial x^i \partial x^k \partial x^l}(x) y^i y^l - \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k}(x) y^i y^j \\ &= \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k}(x) y^i y^j. \end{aligned} \quad (3-119)$$

Substituindo esse resultado na equação (3-117) temos

$$G^i = \frac{1}{4} g^{ik}(x) \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k}(x) y^j y^j. \quad (3-120)$$

Por definição de estrutura afim temos que

$$g_{ij}^*(x) = g_{ij}(x)$$

e

$$H_{ijk}^*(x) = -H_{ijk}(x).$$

sendo assim,

$$H^*(x, y) = H(x, -y) = \frac{1}{6\alpha} \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k}(x) y^i y^j y^k = -H(x, y). \quad (3-121)$$

Assim os coeficientes do α spray G_α^* de (L^*, H^*) é dado por

$$\begin{aligned} G_\alpha^{*i}(x, y) &= \mathcal{G}^i(x, y) - \frac{\alpha}{2} g^{ik} H_{y^k}(x, y) \\ &= \mathcal{G}^i(x, y) - \frac{1}{2} g^{ik} (\alpha H)_{,y^k}(x, y). \end{aligned} \quad (3-122)$$

Utilizando a equação (3-51) temos que

$$\begin{aligned} G_\alpha^{*i}(x, y) &= \mathcal{G}^i(x, y) - \frac{1}{2} g^{ik} \left(\frac{-1}{6} L_{x^l} y^l \right)_{,y^k}(x, y) \\ &= \mathcal{G}^i(x, y) - \frac{1}{2} g^{ik} \left(\frac{-1}{6} L_{x^l, y^k} y^l - \frac{1}{6} L_{x^k} \right)_{,y^k}(x, y). \end{aligned} \quad (3-123)$$

Por hipótese, (L, H) é α -flat, então (3-50) é satisfeita, então:

$$\begin{aligned} G_\alpha^{*i}(x, y) &= \mathcal{G}^i(x, y) - \frac{1}{2} g^{ik} \left[\frac{-1}{6} (2L_{x^k}) - \frac{1}{6} L_{x^k} \right]_{,y^k}(x, y) \\ &= \mathcal{G}^i(x, y) + \frac{1}{4} g^{ik} L_{x^k}. \end{aligned} \quad (3-124)$$

Utilizando a equação (3-122) e a expressão para L_{x^k} , temos

$$\begin{aligned} G_\alpha^{*i}(x, y) &= \frac{1}{4} g^{ik}(x) \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k}(x) y^i y^j + \frac{1}{4} g^{ik} \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k}(x) y^i y^j \\ &= \frac{1}{2} g^{ik} \frac{\partial^3 \Psi}{\partial x^i \partial x^j \partial x^k}(x) y^i y^j. \end{aligned} \quad (3-125)$$

Isto nos diz que, como L^* também é Riemanniana, os símbolos de Christoffel $(\Gamma_\alpha)_{jk}^{*i}$ de G_α^* são dados por

$$(\Gamma_\alpha)_{jk}^{*i}(x) = g^{il}(x) \frac{\partial^3 \Psi}{\partial x^j \partial x^k \partial x^l}(x).$$

Vamos agora determinar um sistema de coordenadas locais (x_i^*) onde os coeficientes do spray G^* sejam nulos. Para isso, considere a seguinte aplicação:

$$x_i^*(x) = \frac{\partial \Psi}{\partial x^i}(x).$$

Podemos observar que o jacobiano de $x^* = x^*(x)$ é justamente a matriz g_{ij} da métrica L , ou seja,

$$g_{ij}(x) = \frac{\partial x_i^*}{\partial x^j} = \frac{\partial^2 \Psi}{\partial x^i \partial x^j}.$$

A aplicação x_i^* é um difeomorfismo local que serve como uma mudança de coordenadas. Definamos agora $\Psi^*(x^*)$ como na equação (3-114). Utilizando agora como hipótese a equação (3-114), chegamos à seguinte conclusão:

$$\begin{aligned} \frac{\partial \Psi^*(x^*)}{\partial x_k^*} &= -\frac{\partial \Psi(x)}{\partial x_k^*} + \sum_{i=1}^n \left[\frac{\partial x_i^*}{\partial x_k^*} x^i + x_i^* \frac{\partial x^i}{\partial x_k^*} \right] \\ &= -\sum_{i=1}^n \frac{\partial \Psi(x)}{\partial x^i} \frac{\partial x^i}{\partial x_k^*} + \delta_{ik} x^i + \sum_{i=1}^n \frac{\partial \Psi(x)}{\partial x^i} \frac{\partial x^i}{\partial x_k^*} \\ &= x^k. \end{aligned} \tag{3-126}$$

Desde que (L^*, H^*) é afim, podemos expressar L^* e H^* em um sistema de coordenadas $(x^*)^i$ como

$$L^* = g^{*kl}(x^*) y_k^* y_l^*$$

e

$$H^* = H^{*ijk}(x^*) y_i^* y_j^* y_k^*.$$

Primeiramente sabemos que

$$x^j = \frac{\partial \Psi^*(x^*)}{\partial x_i^*}.$$

Então, olhando para a equação (3-114) temos que ela assume a seguinte forma:

$$\Psi^*(x^*) := -\Psi(x) + \sum_{i=1}^n x_i^* \frac{\partial \Psi^*(x^*)}{\partial x_i^*}.$$

Já sabemos que o jacobiano da transformação $x^* = x^*(x)$ é a matriz $g_{ij}(x)$, ou seja ,

$$\frac{\partial x_i^*}{\partial x^j} = g_{ij}(x). \quad (3-127)$$

Então para g^{*kl} temos que

$$\begin{aligned} g^{*kl} &= \frac{\partial x^k}{\partial x_l^*} \\ &= \frac{\partial}{\partial x_l^*} \left(\frac{\partial \psi^*(x^*)}{\partial x_k^*} \right) \\ &= \frac{\partial^2 \psi^*(x^*)}{\partial x_l^* \partial x_k^*}. \end{aligned} \quad (3-128)$$

Além disso,

$$\begin{aligned} \frac{\partial x_i^*}{\partial x^j}(x) \cdot (\Gamma \alpha)_{jk}^{*l}(x) &= \frac{\partial x_i^*}{\partial x^j}(x) \cdot g^{ls} \frac{\partial^3 \psi}{\partial x^j \partial x^k \partial x^s} \\ &= g_{il} g^{ls} \frac{\partial^3 \psi}{\partial x^j \partial x^k \partial x^s} \\ &= \delta_{is} \frac{\partial^3 \psi}{\partial x^j \partial x^k \partial x^s} \\ &= \frac{\partial^3 \psi}{\partial x^i \partial x^j \partial x^k}, \end{aligned} \quad (3-129)$$

Também temos que

$$\frac{\partial^2 x_i^*(x)}{\partial x^j \partial x^k} = \frac{\partial}{\partial x^j} \left(\frac{\partial x_i^*}{\partial x^k} \right).$$

Utilizando a equação (3-127) temos que

$$\begin{aligned} \frac{\partial^2 x_i^*(x)}{\partial x^j \partial x^k} &= \frac{\partial g_{ik}}{\partial x^j} \\ &= \frac{\partial}{\partial x^j} \left(\frac{\partial^2 \psi(x)}{\partial x^i \partial x^k} \right) \\ &= \frac{\partial^3 \psi(x)}{\partial x^i \partial x^j \partial x^k}. \end{aligned} \quad (3-130)$$

Considerando as equações (3-127) e (3-130), (3-129)) temos

$$\frac{\partial x_i^*}{\partial x^j}(x) \cdot (\Gamma \alpha)^{*j}_i(x) - \frac{\partial^2 x_i^*(x)}{\partial x^j \partial x^k} = 0. \quad (3-131)$$

Logo, nesse novo sistema de coordenadas, $G_\alpha^{*i} = 0$, assim a estrutura de informação (L^*, H^*) é α -flat. Então podemos escrever H^* como

$$H^*(x^*, y^*) = -\frac{1}{6\alpha} \frac{\partial^3 \Psi^*(x^*)}{\partial x_i^* \partial x_j^* \partial x_k^*} y_i^* y_j^* y_k^*.$$

□

Teorema 3.24 *Seja $\alpha \neq 0$. Uma estrutura de informação (L, H) afim é α -flat se e somente se sua estrutura de informação dual (L^*, H^*) é α -flat.*

Prova.

Se (L, H) é α -flat e afim, então L e H podem ser escritos como nas equações (3-109) e (3-110). Então pelo Lema 3.23, temos que (L^*, H^*) é α -flat. Por outro lado, se (L^*, H^*) é α -flat e como (L, H) é afim então $(L^*, H^*) = (L, -H)$ também é afim. Logo (L^*, H^*) pode também ser escrita na forma das equações (3-109) e (3-110). Daí, pelo Lema 3.23, o seu dual, que é dual de $(L, -H)$ ou seja, (L, H) é α -flat.

□

3.8 Conexões Afins Dualísticas

Sabemos que as conexões afins têm correspondência 1 – 1 com os sprays afins. Uma conexão afim em uma Variedade Riemanniana (M, g) é chamada de **dualística** se a conexão linear dual ∇^* com respeito a g é afim. Seja $L = g_{ij}(x)y^i y^j$ uma Métrica Riemanniana em M e $g = g_{ij} dx^i \otimes dx^j$ o produto interno associado em espaços tangentes. Para uma conexão ∇ em M defina ∇^* como se segue:

$$g(\nabla_Z^* X, Y) + g(X, \nabla_Z Y) = Z(g(X, Y)), \quad (3-132)$$

onde $X, Y, Z \in C^\infty(TM)$. Temos que ∇^* também é linear.

Um importante fenômeno é que se a conexão linear ∇ é afim, a conexão linear dual não é necessariamente afim, isto é, não tem torção nula.

Teorema 3.25 *Seja g uma métrica Riemanianna em uma variedade M . Toda H -função polinomial em (M, g) determina uma conexão afim dualística. Inversamente toda conexão*

afim dualística ∇ determina uma H -função. A correspondência é canônica:

$$\Gamma_{jk}^i = \Upsilon_{jk}^i + 3g^{il}H_{jkl}(x), \quad (3-133)$$

onde Γ_{jk}^i são os símbolos de Christoffel de ∇ e Υ_{jk}^i denota os símbolos de Christoffel de g .

Prova.

Seja H uma H -função polinomial em uma variedade Riemanniana (M, g) . Sejam ∇ e $\bar{\nabla}$ conexões afins correspondentes aos 1-sprays associados G_1 e \bar{G}_1 das estruturas de informação (g, H) e $(g, -H)$, respectivamente. Notemos que $(g, -H)$ é a informação dual de (g, H) . Afiramos que

$$g(\bar{\nabla}_Z X, Y) + g(X, \nabla_Z Y) = Z(g(X, Y)). \quad (3-134)$$

Ou seja, $\bar{\nabla}$ é dual a ∇ com respeito a g . Seja $g = g_{ij}(x)y^i y^j$ e $H = H_{ijk}y^i y^j y^k$. Sejam $\Gamma_{jk}^i(x)$ e $\bar{\Gamma}_{jk}^i(x)$ os símbolos de christoffel de G_1 e \bar{G}_1 respectivamente. Sejam também $\Gamma_{jk,i}(x) = g_{il}\Gamma_{jk}^l(x)$ e $\bar{\Gamma}_{jk,i}(x) = g_{il}\bar{\Gamma}_{jk}^l(x)$. Sabemos que

$$G_1^i(x, y) = \mathcal{G}^i(x, y) + \frac{g^{ij}(x)}{2}H_{y^j}(x, y), \quad (3-135)$$

como $H = H_{ijk}y^i y^j y^k$, temos que

$$H_{y^j} = 3H_{ijk}(x)y^i y^k. \quad (3-136)$$

Substituindo esse resultado na equação (3-135), levando em conta que g é uma métrica riemanniana, logo $\mathcal{G}^l(x, y) = \Upsilon_{jk}^l(x)y^j y^k$ e que G_1 é afim,

$$\frac{1}{2}\Gamma_{jk}^l(x)y^j y^k = G_1^l(x, y) = \frac{1}{2}\Upsilon_{jk}^l(x)y^j y^k + \frac{3}{2}g^{il}(x)H_{ijk}(x)y^j y^k,$$

de onde temos

$$\Gamma_{jk}^l(x) = \Upsilon_{jk}^l(x) + 3g^{lr}(x)H_{rjk}(x).$$

Multiplicando essa equação por g_{il} temos

$$g_{il}\Gamma_{jk}^l(x) = g_{il}\Upsilon_{jk}^l(x) + 3\delta_{ir}(x)H_{rjk}(x).$$

O que resulta em

$$\Gamma_{jk,i}(x) = \Upsilon_{jk,i}(x) + 3H_{ijk}(x) \quad (3-137)$$

Analogamente para \bar{G}_1 temos

$$\frac{1}{2}\bar{\Gamma}_{ik}^l y^i y^k = \frac{1}{2}\bar{\gamma}_{ik}^l y^i y^k - \frac{3}{2}g^{ij}H_{ijk}(x)y^i y^k.$$

Ao multiplicar toda essa equação por g_{jl} resulta-se em

$$\bar{\Gamma}_{ik,j}(x) = \gamma_{ik,j}(x) - 3H_{ijk}(x). \quad (3-138)$$

Somando as equações (3-137) e (3-138) temos

$$\begin{aligned} \bar{\Gamma}_{ik,j}(x) + \Gamma_{jk,i}(x) &= \gamma_{ik,j}(x) + \gamma_{jk,i}(x) \\ &= g_{jk}\gamma_{ik}^l + g_{il}\gamma_{jk}^l \\ &= g_{jl} \left[\frac{1}{2}g^{ls} \left(\frac{\partial g_{is}}{\partial x^k} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ik}}{\partial x^s} \right) \right] + g_{il} \left[\frac{1}{2}g^{ls} \left(\frac{\partial g_{js}}{\partial x^k} + \frac{\partial g_{ks}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^s} \right) \right] \\ &= \frac{1}{2}\delta_{js} \left(\frac{\partial g_{is}}{\partial x^k} + \frac{\partial g_{kj}}{\partial x^i} - \frac{\partial g_{ik}}{\partial x^s} \right) + \frac{1}{2}\delta_{is} \left(\frac{\partial g_{js}}{\partial x^k} + \frac{\partial g_{ks}}{\partial x^j} - \frac{\partial g_{jk}}{\partial x^s} \right) \\ &= \frac{1}{2} \frac{\partial g_{ij}}{\partial x^k} + \frac{1}{2} \frac{\partial g_{ij}}{\partial x^k} \\ &= \frac{\partial g_{ij}}{\partial x^k}. \end{aligned} \quad (3-139)$$

Mas sabemos que

$$g(\nabla_{X_i} X_j, X_k) = \Gamma_{ij}^l g_{kl} = \Gamma_{ij,k},$$

e de maneira analoga,

$$g(\bar{\nabla}_{X_i} X_k, X_j) = \bar{\Gamma}_{ik}^l g_{jl} = \Gamma_{ik,j}.$$

sendo assim

$$g(\nabla_{X_i} X_j, X_k) + g(\bar{\nabla}_{X_i} X_k, X_j) = \frac{\partial g(X_i, X_j)}{\partial x^k}. \quad (3-140)$$

Assim, escrevemos (3-140) como (3-134). Sendo assim, $\bar{\nabla} = \nabla^*$, ou seja: $\bar{\nabla}$ é a conexão linear dual de ∇ em (M, g) . Como $\bar{\nabla}$ é afim, temos, por definição, que ∇ é dualística. Seja ∇ uma conexão afim em (M, g) . Defina H_{ijk} como em (3-137). Temos então que $H_{ijk}(x) = H_{ikj}(x)$ pois obviamente $\Gamma_{jk,i} = \Gamma_{k,j,i}$ e $\gamma_{jk,i} = \gamma_{k,j,i}$. Seja ∇^* a conexão linear dual. Seja Γ_{jk}^{*i} os Símbolos de Christoffel de ∇^* e $\Gamma_{jk,l}^* = g_{il}\Gamma_{jk}^{*i}$. Então

$$\Gamma_{ik,j}^*(x) + \Gamma_{jk,i}(x) = \frac{\partial g_{ij}(x)}{\partial x^k} = \gamma_{ik,j}(x) + \Gamma_{jk,i}(x). \quad (3-141)$$

Utilizando a equação (3-137), temos que

$$\begin{aligned}\Upsilon_{ik,j}(x) + \Upsilon_{jk,i}(x) &= \Gamma_{ik,j}^*(x) + \Gamma_{jk,i}(x) \\ &= \Gamma_{ik,j}^*(x) + \Upsilon_{jk,i}(x) + 3H_{ijk}(x).\end{aligned}\tag{3-142}$$

Logo

$$\Gamma_{ik,j}^*(x) = \Upsilon_{ik,j}(x) - 3H_{ijk}(x).$$

Assim

$$3H_{ijk}(x) = \Upsilon_{ik,j}(x) - \Gamma_{ik,j}^*(x).$$

Podemos permutar i e k logo

$$H_{ijk}(x) = H_{kji}(x).$$

Logo H é simétrico em relação a i, j e k . Então obtemos uma H -função $H = H_{ijk}(x)y^i y^j y^k$. Vamos reescrever (3-137) como

$$\Gamma_{jk,l}(x) = \Upsilon_{jk,l}(x) + 3H_{ljk}(x).$$

Temos então

$$g_{sl}\Gamma_{jk}^s(x) = g_{sl}\Upsilon_{jk}^s(x) + 3H_{ljk}(x).$$

Multiplicando toda essa equação por g^{il} , obtemos que

$$\Gamma_{jk}^i(x) = \Upsilon_{jk}^i(x) + 3g^{il}H_{ljk}(x).$$

Utilizando a simetria de H , temos finalmente

$$\Gamma_{jk}^i(x) = \Upsilon_{jk}^i(x) + 3g^{il}H_{jkl}(x).$$

□

Teorema 3.26 (Amari) *Sejam ∇ e ∇^* as conexões duais afins em uma variedade Riemanniana (M, g) . Então ∇ é flat se, e somente se, ∇^* é flat.*

Prova. Seja H a H -função polinomial correspondente a ∇ . Então $H^* = -H$ é a H -função correspondente a ∇^* . Notemos que o spray associado a (g, H) tem correspondência biunívoca com ∇ , e que o spray associado a (g, H^*) tem correspondência biunívoca com ∇^* . Logo ∇ é flat se e somente se, (g, H) é 1-flat. Mas, segundo o teorema 3.24 (g, H) é

1-flat se, e somente se, (g, H^*) é 1-flat. Mas (g, H^*) é 1-flat se, e somente se, ∇^* é flat. Logo, ∇ é flat se, e somente se, ∇^* é flat. \square

Aplicações aos Modelos estatísticos

4.1 f -divergências em espaços de probabilidade

Seja $\mathcal{X} = (\mathcal{X}, \mathcal{B}, \mathcal{V})$ um espaço de medida onde \mathcal{X} é um conjunto não vazio, \mathcal{B} é uma classe aditiva completa que consiste em \mathcal{X} e seus subconjuntos, e \mathcal{V} uma medida σ -finita em $(\mathcal{X}, \mathcal{B})$. Seja $\mathcal{P} = \mathcal{P}(\mathcal{X})$ o espaço das densidades de probabilidade em \mathcal{X} , ou seja,

$$\mathcal{P}(\mathcal{X}) = \{p : \mathcal{X} \rightarrow [0, \infty) \mid \int_{\mathcal{X}} p(r) dr = 1\}.$$

Temos que \mathcal{P} é um conjunto convexo. De fato, sejam $p, q : \mathcal{X} \rightarrow [0, \infty) \in \mathcal{P}$ e $\lambda \in [0, 1]$. Defina $\alpha : \mathcal{X} \rightarrow [0, \infty)$, $\alpha = \lambda p + (1 - \lambda)q$. Então

$$\begin{aligned} \int_{\mathcal{X}} \alpha dr &= \int_{\mathcal{X}} \lambda p + (1 - \lambda)q dr \\ &= \lambda \underbrace{\int_{\mathcal{X}} p dr}_1 + (1 - \lambda) \underbrace{\int_{\mathcal{X}} q dr}_1 \\ &= \lambda + (1 - \lambda) \\ &= 1. \end{aligned} \tag{4-1}$$

Assim $\alpha \in \mathcal{P}$, logo \mathcal{P} é convexo.

Seja f uma função convexa $f : (0, \infty) \rightarrow \mathbb{R}$ com $f(1) = 0$ e $f''(1) = 1$. Defina

$$D_f : \mathcal{P} \times \mathcal{P} \rightarrow \mathbb{R}$$

$$D_f(p, q) = \int_{\mathcal{X}} p(r) f\left(\frac{q(r)}{p(r)}\right) dr \tag{4-2}$$

$p, q \in \mathcal{P}$. Temos que D_f é uma divergência em \mathcal{P} . O que de fato é verdade, pois como f

é contínua e $p \geq 0$, utilizando a Desigualdade de Jensen, temos que

$$\begin{aligned} D_f(p, q) &= \int_X p(r) f\left(\frac{q(r)}{p(r)}\right) dr \\ &\geq f\left(\int_X p(r) \frac{q(r)}{p(r)} dr\right) \\ &= f\left(\int_X q(r) dr\right) \\ &= f(1) \\ &= 0, \end{aligned}$$

(4-3)

onde a igualdade ocorre se e somente se $p = q$. Vamos agora construir uma família especial de f -divergências em \mathcal{P} . Para $\rho \in \mathbb{R}$ seja

$$f_\rho(t) = \begin{cases} \frac{4}{1-\rho^2} \left(\frac{1+t}{2} - t^{\frac{1+\rho}{2}} \right) & \text{se } \rho \neq \pm 1 \\ t \ln(t) & \text{se } \rho = 1 \\ -\ln(t) & \text{se } \rho = -1 \end{cases}$$

Então

$$\begin{aligned} f'_\rho(t) &= \begin{cases} \frac{4}{1-\rho^2} \left(\frac{1}{2} - \frac{1+\rho}{2} t^{\frac{1+\rho}{2}-1} \right) & \text{se } \rho \neq \pm 1 \\ 1 + \ln(t) & \text{se } \rho = 1 \\ -\frac{1}{t} & \text{se } \rho = -1 \end{cases} \\ f''_\rho(t) &= \begin{cases} \frac{4}{1-\rho^2} \left(\left(\frac{1+\rho}{2} \right) \left(\frac{1-\rho}{2} \right) t^{\frac{1+\rho}{2}-2} \right) & \text{se } \rho \neq \pm 1 \\ \frac{1}{t} & \text{se } \rho = 1 \\ \frac{1}{t^2} & \text{se } \rho = -1 \end{cases} \end{aligned}$$

e $f''_\rho(1) = 1$. Para $\rho = 0$ temos

$$f_0(t) = 4 \left(\frac{1+t}{2} - \sqrt{t} \right) = 2(1+t) - 4\sqrt{t}.$$

A divergência D_0 em \mathcal{P} é dada por

$$\begin{aligned}
 D_0(p, q) &= \int_X p(r) f\left(\frac{q(r)}{p(r)}\right) dr \\
 &= \int_X p(r) \left[2\left(1 + \frac{q(r)}{p(r)}\right) - 4\sqrt{\frac{q(r)}{p(r)}} \right] dr \\
 &= 2 \int_X (p(r) + q(r)) dr - \int_X 4p(r) \sqrt{\frac{q(r)p(r)}{p(r)^2}} dr \\
 &= 2 \int_X (p(r) + q(r)) dr - \int_X 4\sqrt{p(r)q(r)} dr \\
 &= 2 \int_X p(r) + q(r) - 2\sqrt{p(r)q(r)} dr \\
 &= 2 \int_X \left(\sqrt{p(r)} - \sqrt{q(r)} \right)^2 dr.
 \end{aligned} \tag{4-4}$$

Podemos ver que $d_0(p, q) = \sqrt{D_0(p, q)}$ é uma função distância pois $d_0 \geq 0$ e $d_0(p, q) = 0$ se, e somente se, $D_0(p, q) = 0 \Leftrightarrow p = q$ e

$$d_0(p, s) + d_0(s, q) = \left[2 \left(\int_X (\sqrt{p} - \sqrt{s})^2 ds \right)^{\frac{1}{2}} + 2 \left(\int_X (\sqrt{s} - \sqrt{q})^2 ds \right)^{\frac{1}{2}} \right].$$

A **desigualdade de Minkowski** nos diz que se $f, g : \Omega \rightarrow \mathbb{R}$ são funções contínuas, então

$$\left(\int_{\Omega} \|f + g\|^p \right)^{\frac{1}{p}} \leq \left(\int_{\Omega} \|f\|^p \right)^{\frac{1}{p}} + \left(\int_{\Omega} \|g\|^p \right)^{\frac{1}{p}}. \tag{4-5}$$

então, utilizando (3.5), temos

$$\begin{aligned}
 d_0(p, s) + d_0(s, q) &\geq 2 \left(\int_X (\sqrt{p} - \sqrt{s} + \sqrt{s} - \sqrt{q})^2 \right)^{\frac{1}{2}} \\
 &= \sqrt{2.2 \int_X (\sqrt{p} - \sqrt{q})^2 dr} \\
 &= \sqrt{2D_0(p, q)} \\
 &= d_0(p, q).
 \end{aligned} \tag{4-6}$$

A distância $d_0(p, q)$ é chamada de *distância de Hellinger* e $D_0 := \frac{1}{2}d_0^2$ é chamada

de *Divergência de Hellinger*. Para $\rho = -1$ temos que

$$D_{-1}(p, q) = \int_{\mathcal{X}} p(r) \ln \left(\frac{p(r)}{q(r)} \right) dr,$$

que é chamada de *divergência de Kullback-Leiber*. para $\rho = 1$ temos que

$$D(p, q) = \int_{\mathcal{X}} p(r) \frac{q(r)}{p(r)} \ln \left(\frac{q(r)}{p(r)} \right) dr,$$

logo

$$D(p, q) = \int_{\mathcal{X}} q(r) \ln \left(\frac{q(r)}{p(r)} \right) dr.$$

4.2 Os α -sprays em Modelos Estatísticos

Seja \mathcal{P} um espaço de densidades de probabilidade em um espaço mensurável \mathcal{X} e \mathcal{D} uma divergência em \mathcal{P} . Um modelo estatístico em $(\mathcal{P}, \mathcal{D})$ é um par (M, D) , onde M é uma variedade C^∞ mergulhada em \mathcal{P} e $D = \mathcal{D}|_M$. Se f é uma função $f: (0, \infty) \rightarrow \mathbb{R}$ convexa, tal que $f(1) = 0$ e $f''(1) = 1$, então podemos definir a f -divergência D_f em \mathcal{P} como em (4-2). Agora, vamos provar que para toda variedade $M \subset \mathcal{P}$, a divergência induzida $D_f = \mathcal{D}_{f|M}$ é afim, a saber, a métrica induzida $L = g_{ij}(x)y^i y^j$ é riemanniana e a H -função induzida é polinomial.

Teorema 4.1 *Seja $f = f(t)$ uma função $f(1) = 0$ e $f''(1) = 1$. Para todo modelo estatístico regular (M, D_f) de $(\mathcal{P}, \mathcal{D}_f)$, a estrutura de informação induzida em M é dada por $(L_f, H_f) = (L, \rho N)$, onde $\rho := 3 + f'''(1)$ e*

$$L = \int_{\mathcal{X}} \left[y^i \frac{\partial}{\partial x^i} \ln p \right]^2 p dr = \mathbb{E} \left[\left(\frac{\partial \ln p(r; x)}{\partial x^i} y^i \right)^2 \right], \quad (4-7)$$

$$N = \frac{1}{6} \int_{\mathcal{X}} \left[y^i \frac{\partial}{\partial x^i} \ln p \right]^3 p dr = \mathbb{E} \left[\frac{1}{6} \left(\frac{\partial \ln p(r; x)}{\partial x^i} y^i \right)^3 \right], \quad (4-8)$$

onde \mathbb{E} denota a esperança matemática. O α -spray $G_{\alpha, \rho}$ de D_f é dado por

$$G_{\alpha, \rho}^i = \bar{G}^i + (\rho \alpha + 1) A^i,$$

onde

$$\bar{G}^i = \frac{g^{il}(x)}{2} \int_{\mathcal{X}} \left[y^j y^j \frac{\partial^2}{\partial x^i \partial x^j} \ln p(r; x) \right] \frac{\partial}{\partial x^l} p(r; x) dr \quad (4-9)$$

e

$$A^i = \frac{g^{il}(x)}{4} \int_{\mathcal{X}} \left[y^j \frac{\partial}{\partial x^j} \ln p(r; x) \right]^2 \frac{\partial}{\partial x^i} p(r; x) dr. \quad (4-10)$$

Prova. O mergulho natural $M \rightarrow \mathcal{P}$ é o que leva um ponto x de M em um ponto p de \mathcal{P} , ou seja, $x \rightarrow p = p(r; x)$. Seja $D(x, z) := D_f(p(r; x), p(r; z))$, isto é,

$$D(x, z) := \int_X p(r; x) f\left(\frac{p(r; z)}{p(r; x)}\right) dr. \quad (4-11)$$

Expandindo D , temos que

$$2D(x, x+y) = \frac{\partial^2 D}{\partial z^i \partial z^j} \Big|_{z=x} y^i y^j + \frac{1}{3} \frac{\partial^3 D}{\partial z^i \partial z^j \partial z^k} \Big|_{z=x} y^i y^j y^k + o(|y|^3), \quad (4-12)$$

onde

$$\begin{aligned} D|_{z=x} &= \int_X p(r; x) f\left(\frac{p(r; x)}{p(r; x)}\right) dr \\ &= \int_X p(r; x) f(1) dr \\ &= 0 \end{aligned}$$

Seja $h = \frac{p(r; z)}{p(r; x)}$. Temos então que

$$\begin{aligned} \frac{\partial D}{\partial z^i} &= \int_X p(r; x) \frac{\partial h}{\partial z^i} \frac{df(h)}{dh} dr \\ &= \int_X p(r; x) \frac{1}{p(r; x)} \frac{\partial p(r; z)}{\partial z^i} \frac{df(h)}{dh} dr \\ &= \int_X \frac{\partial p(r; z)}{\partial z^i} \frac{df(h)}{dh} dr. \end{aligned} \quad (4-13)$$

Utilizando a restrição $z = x$, temos que $f'(h) = f'(1)$, então

$$\begin{aligned} \frac{\partial D}{\partial z^i} \Big|_{z=x} y^i &= f'(1) \frac{\partial}{\partial x^i} \int_X p(r; x) y^i dr \\ &= f'(1) \frac{\partial(1)}{\partial x^i} y^i \\ &= 0. \end{aligned}$$

Temos que

$$\frac{\partial D}{\partial z^j} y^j = \int_X \frac{\partial p(r; z)}{\partial z^j} \frac{df(h)}{dh} y^j dr,$$

derivando-se essa expressão em relação a z^i e somando-se em y^i temos

$$\frac{\partial^2 D}{\partial z^i \partial z^j} y^i y^j = \int_X \frac{\partial^2 p(r; z)}{\partial z^i \partial z^j} \frac{df(h)}{dh} y^i y^j dr +$$

Prova. O mergulho natural $M \rightarrow \mathcal{P}$ é o que leva um ponto x de M em um ponto p de \mathcal{P} , ou seja, $x \rightarrow p = p(r; x)$. Seja $D(x, z) := D_f(p(r; x), p(r; z))$, isto é,

$$D(x, z) := \int_X p(r; x) f\left(\frac{p(r; z)}{p(r; x)}\right) dr. \quad (4-11)$$

Expandindo D , temos que

$$2D(x, x+y) = \frac{\partial^2 D}{\partial z^i \partial z^j} \Big|_{z=x} y^i y^j + \frac{1}{3} \frac{\partial^3 D}{\partial z^i \partial z^j \partial z^k} \Big|_{z=x} y^i y^j y^k + o(|y|^3), \quad (4-12)$$

onde

$$\begin{aligned} D|_{z=x} &= \int_X p(r; x) f\left(\frac{p(r; x)}{p(r; x)}\right) dr \\ &= \int_X p(r; x) f(1) dr \\ &= 0 \end{aligned}$$

Seja $h = \frac{p(r; z)}{p(r; x)}$. Temos então que

$$\begin{aligned} \frac{\partial D}{\partial z^i} &= \int_X p(r; x) \frac{\partial h}{\partial z^i} \frac{df(h)}{dh} dr \\ &= \int_X p(r; x) \frac{1}{p(r; x)} \frac{\partial p(r; z)}{\partial z^i} \frac{df(h)}{dh} dr \\ &= \int_X \frac{\partial p(r; z)}{\partial z^i} \frac{df(h)}{dh} dr. \end{aligned} \quad (4-13)$$

Utilizando a restrição $z = x$, temos que $f'(h) = f'(1)$, então

$$\begin{aligned} \frac{\partial D}{\partial z^i} \Big|_{z=x} y^i &= f'(1) \frac{\partial}{\partial x^i} \int_X p(r; x) y^i dr \\ &= f'(1) \frac{\partial(1)}{\partial x^i} y^i \\ &= 0. \end{aligned}$$

Temos que

$$\frac{\partial D}{\partial z^j} y^j = \int_X \frac{\partial p(r; z)}{\partial z^j} \frac{df(h)}{dh} y^j dr,$$

derivando-se essa expressão em relação a z^i e somando-se em y^i temos

$$\frac{\partial^2 D}{\partial z^i \partial z^j} y^i y^j = \int_X \frac{\partial^2 p(r; z)}{\partial z^i \partial z^j} \frac{df(h)}{dh} y^i y^j dr +$$

$$+ \int_X \frac{\partial^2 p(r; z)}{\partial z^j \partial z^k} \frac{d^2 f(h)}{dh^2} \frac{\partial h}{\partial z^i} y^j y^k dr.$$

Assim, essa expressão fica da forma

$$\begin{aligned} \frac{\partial I(z, x)}{\partial z^i} y^j y^k &= \int_X \frac{\partial^3 p(r; z)}{\partial z^i \partial z^j \partial z^k} \frac{df(h)}{dh} y^j y^k dr + \\ &+ \int_X \frac{\partial^2 p(r; z)}{\partial z^j \partial z^k} \frac{d^2 f(h)}{dh^2} \frac{1}{p(r; x)} \frac{\partial p(r; z)}{\partial z^i} y^j y^k dr. \end{aligned} \quad (4-18)$$

Olhando agora (3-139) com a restrição $z = x$ temos

$$\begin{aligned} \frac{\partial I(z, x)}{\partial z^i} \Big|_{z=x} y^j y^k &= f'(1) \frac{\partial^3}{\partial x^i \partial x^j \partial x^k} \int_X p(r; x) y^j y^k dr + \\ &+ f''(1) \int_X \frac{\partial^2 p(r; z)}{\partial z^j \partial z^k} \frac{1}{p(r; x)} \frac{\partial p(r; z)}{\partial z^i} y^j y^k dr. \end{aligned} \quad (4-19)$$

Então

$$\frac{\partial I(z, x)}{\partial z^i} \Big|_{z=x} y^j y^k = \int_X \frac{\partial^2 p(r; z)}{\partial x^j \partial x^k} \frac{1}{p(r; x)} \frac{\partial p(r; x)}{\partial x^i} y^j y^k dr. \quad (4-20)$$

Utilizando o fato (4-16) temos finalmente que

$$\frac{\partial I(z, x)}{\partial z^i} \Big|_{z=x} y^j y^k = \int_X \frac{\partial^2 p(r; z)}{\partial x^j \partial x^k} \frac{\partial \ln p(r; x)}{\partial x^i} y^j y^k dr. \quad (4-21)$$

Vamos derivar $II(z, x)$ em relação a z^i e somar em y^j , assim, teremos

$$\frac{\partial II(z, x)}{\partial z^i} y^j y^k = \underbrace{\int_X \frac{\partial^2 p(r; z)}{\partial z^i \partial z^k} \frac{1}{p(r; x)} \frac{\partial p(r; z)}{\partial z^j} \frac{d^2 f(h)}{dh^2} y^j y^k}_{a(z, x)} + \quad (4-22)$$

$$+ \underbrace{\int_X \frac{\partial p(r; z)}{\partial z^k} \frac{1}{p(r; x)} \frac{\partial^2 p(r; z)}{\partial z^j \partial z^i} \frac{d^2 f(h)}{dh^2} y^j y^k}_{b(z, x)} + \quad (4-23)$$

$$+ \underbrace{\int_X \frac{\partial p(r; z)}{\partial z^k} \frac{1}{p(r; x)} \frac{\partial p(r; z)}{\partial z^j} \frac{1}{p(r; x)} \frac{\partial h}{\partial z^i} \frac{d^3 f(h)}{dh^3} y^j y^k}_{c(z, x)}. \quad (4-24)$$

Lembrando-se que $\frac{\partial h}{\partial z^i} = \frac{1}{p(r;x)} \frac{\partial p(r;x)}{\partial z^i}$ e olhando essa expressão na restrição $z = x$, temos então que

$$\begin{aligned} & a(z,x)|_{z=x} y^i y^j y^k + b(z,x)|_{z=x} y^i y^j y^k + \frac{\partial I(z,x)}{\partial z^i} \Big|_{z=x} y^i y^j y^k = \quad (4-25) \\ & = f'''(1) \left[\int_X \frac{\partial^2 p(r;z)}{\partial x^i \partial x^k} y^i y^j y^k \frac{\partial \ln p(r;x)}{\partial x^j} y^j dr + \int_X \frac{\partial^2 p(r;z)}{\partial x^i \partial x^j} y^i y^j \frac{\partial \ln p(r;x)}{\partial x^k} y^k \right] dr + \\ & \quad + \left[\int_X \frac{\partial^2 p(x;z)}{\partial x^j \partial x^k} y^j y^k \frac{\partial \ln p(r;x)}{\partial x^i} y^i \right] dr. \end{aligned}$$

Assim

$$\begin{aligned} & a(z,x)|_{z=x} y^i y^j y^k + b(z,x)|_{z=x} y^i y^j y^k + \frac{\partial I(z,x)}{\partial z^i} \Big|_{z=x} y^i y^j y^k = \quad (4-26) \\ & = 3 \left[\int_X \frac{\partial^2 p(x;z)}{\partial x^i \partial x^j} y^i y^j \frac{\partial \ln p(r;x)}{\partial x^k} y^k \right] dr. \end{aligned}$$

Olhando $c(x,z)$ dado em (4-24) na restrição $z = x$ e levando em conta (4-16), temos

$$\begin{aligned} c(x,z)|_{z=x} y^i y^j y^k & = f'''(1) \int_X \frac{\partial p(r;x)}{\partial x^k} y^k \frac{1}{p(r;x)} \frac{\partial p(r;x)}{\partial x^j} y^j \frac{1}{p(r;x)} \frac{\partial p(r;x)}{\partial x^i} y^i \frac{1}{p(r;x)} y^i p(r;x) dr \\ & = \frac{\rho-3}{2} \int_X \left[\frac{\partial \ln p(r;x)}{\partial x^i} y^i \right]^3 p(r;x) dr. \end{aligned}$$

Logo

$$\begin{aligned} & \frac{\partial^3 D}{\partial z^i \partial z^j \partial z^k} \Big|_{z=x} y^i y^j y^k = \quad (4-27) \\ & = \frac{\rho}{2} \int_X \left[\frac{\partial \ln p(r;x)}{\partial x^i} y^i \right]^3 p dr + \frac{3}{2} \left\{ - \int_X \left[\frac{\partial \ln p(r;x)}{\partial x^i} y^i \right]^3 p dr + 2 \left[\int_X \frac{\partial^2 p(x;z)}{\partial x^i \partial x^j} y^i y^j \frac{\partial \ln p(r;x)}{\partial x^k} y^k \right] \right\}. \end{aligned}$$

Seja

$$L = \int_X \left[\frac{\partial \ln p(r;x)}{\partial x^i} y^i \right]^2 p dr = \int_X \left[\frac{\partial \ln p(r;x)}{\partial x^i} y^i \right] \left[\frac{\partial \ln p(r;x)}{\partial x^j} y^j \right] p(r;x) dr, \quad (4-28)$$

então derivando-se L em relação a x^k e somando-se em y^k temos

$$L_{,x} y^k = \int_X y^i \frac{\partial^2 \ln p(r;x)}{\partial x^k \partial x^i} y^j \frac{\partial \ln p(r;x)}{\partial x^j} p(r;x) y^k dr +$$

$$\begin{aligned}
& + \int_{\mathcal{X}} y^i \frac{\partial \ln p(r; x)}{\partial x^i} y^j \frac{\partial^2 \ln p(r; x)}{\partial x^k \partial x^j} p(r; x) y^k dr + \\
& + \int_{\mathcal{X}} y^i \frac{\partial \ln p(r; x)}{\partial x^i} y^j \frac{\partial \ln p(r; x)}{\partial x^j} \underbrace{\frac{\partial p(r; x)}{\partial x^k}}_{\frac{\partial \ln p(r; x)}{\partial x^k} p(r; x)} y^k dr.
\end{aligned}$$

Logo

$$L_{x^k} y^k = 2 \int_{\mathcal{X}} \left[y^k \frac{\partial \ln p(r; x)}{\partial x^k} \right] \left[y^i y^j \frac{\partial^2 \ln p(r; x)}{\partial x^i \partial x^j} \right] p(r; x) dr + \int_{\mathcal{X}} \left[\frac{\partial \ln p(r; x)}{\partial x^i} y^i \right]^3 p(r; x) dr.$$

Mas observemos que

$$\begin{aligned}
& L_{x^k} y^k + \int_{\mathcal{X}} \left[\frac{\partial \ln p(r; x)}{\partial x^i} y^i \right]^3 p(r; x) dr + 2 \int_{\mathcal{X}} y^i y^j \frac{\partial^2 p(r; x)}{\partial x^i \partial x^j} y^k \frac{\partial \ln p(r; x)}{\partial x^k} dr = \\
& = 2 \int_{\mathcal{X}} \left[\frac{\partial \ln p(r; x)}{\partial x^i} y^i \right]^3 p(r; x) dr + 2 \int_{\mathcal{X}} \left[y^k \frac{\partial \ln p(r; x)}{\partial x^k} \right] \left[y^i y^j \frac{\partial^2 \ln p(r; x)}{\partial x^i \partial x^j} \cdot p(r; x) - y^i y^j \frac{\partial^2 p(r; x)}{\partial x^i \partial x^j} y^k \right] dr.
\end{aligned} \tag{4-29}$$

Mas

$$\begin{aligned}
y^i y^j \frac{\partial^2 p(r; x)}{\partial x^i \partial x^j} & = y^i \frac{\partial}{\partial x^i} \left[y^j \frac{\partial p(r; x)}{\partial x^j} \right] \\
& = y^i \frac{\partial}{\partial x^i} \left[y^j p(r; x) \frac{\partial \ln p(r; x)}{\partial x^j} \right] \\
& = y^i \frac{\partial p(r; x)}{\partial x^i} y^j \frac{\partial \ln p(r; x)}{\partial x^j} + y^i y^j \frac{\partial^2 \ln p(r; x)}{\partial x^i \partial x^j} p(r; x) \\
& = y^i \frac{\partial \ln p(r; x)}{\partial x^i} y^j \frac{\partial \ln p(r; x)}{\partial x^j} p(r; x) + y^i y^j \frac{\partial^2 \ln p(r; x)}{\partial x^i \partial x^j} p(r; x).
\end{aligned}$$

Logo

$$y^i y^j \frac{\partial^2 \ln p(r; x)}{\partial x^i \partial x^j} \cdot p(r; x) - y^i y^j \frac{\partial^2 p(r; x)}{\partial x^i \partial x^j} y^k = - \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right]^2 p(r; x). \tag{4-30}$$

Então o segundo termo da equação (4-29) se anula e temos

$$L_{x^k} y^k = - \int_{\mathcal{X}} \left[\frac{\partial \ln p(r; x)}{\partial x^i} y^i \right]^3 p(r; x) dr + 2 \int_{\mathcal{X}} y^i y^j \frac{\partial^2 p(r; x)}{\partial x^i \partial x^j} y^k \frac{\partial \ln p(r; x)}{\partial x^k} dr. \tag{4-31}$$

Fazendo

$$6N = \int_X \left[y^i \frac{\partial \ln p(r;x)}{\partial x^i} \right]^3 p(r;x) dr = \mathbb{E} \left[\left(y^i \frac{\partial \ln p(r;x)}{\partial x^i} \right)^3 \right]. \quad (4-32)$$

Logo

$$\begin{aligned} \frac{\partial^3 D}{\partial z^i \partial z^j \partial z^k} \Big|_{z=x} y^i y^j y^k &= \frac{\rho}{2} \cdot 6N + \frac{3}{2} L_{x^k} y^k \\ &= 3\rho N + \frac{3}{2} L_{x^k} y^k. \end{aligned}$$

Sendo assim,

$$\begin{aligned} 2D(x, x+y) &= L(x, y) + \frac{1}{3} \left[3\rho N(x, y) + \frac{3}{2} L_{x^k} y^k \right] + o(|y|^3) \\ &= L(x, y) + \rho N + \frac{1}{2} L_{x^k} y^k + o(|y|^3) \\ &= \mathbb{E} \left[\left(y^i \frac{\partial \ln p(r;x)}{\partial x^i} \right)^2 \right] + \rho \mathbb{E} \left[\frac{1}{6} \left(y^i \frac{\partial \ln p(r;x)}{\partial x^i} \right)^3 \right] + \frac{1}{2} L_{x^k} y^k + o(|y|^3). \end{aligned}$$

Portanto D_f é uma divergência regular e a estrutura de informação induzida $(L_f, H_f) = (L, \rho N)$ é afim.

Seja $\mathcal{G} = y^i \frac{\partial}{\partial x^i} - 2G^i \frac{\partial}{\partial y^i}$ o spray induzido de L e $G_{\alpha, f} = y^i \frac{\partial}{\partial x^i} - 2G_{\alpha, \rho}^i \frac{\partial}{\partial y^i}$ o α -spray de D_f . Vamos agora obter a expressão para os coeficientes desse α -spray. Sabemos que a expressão para um α -spray de uma estrutura (L, H) é dado por

$$G_{\alpha}^i(x, y) = G^i(x, y) + \frac{\alpha}{2} g^{il}(x, y) H_{y^l}(x, y).$$

Reescrevendo esta expressão para a estrutura de informação $(L, \rho N)$, temos

$$G_{\alpha, \rho}^i(x, y) = G^i(x, y) + \frac{\alpha \rho}{2} g^{il}(x, y) N_{y^l}(x, y) \quad (4-33)$$

onde

$$G^i(x, y) = \frac{1}{4} g^{il}(x, y) \{ L_{y^l x^k} y^k - L_{x^l} \}. \quad (4-34)$$

Temos que L é dada por (4-28). Derivando L em relação a y^l teremos

$$L_{y^l} = \int_X 2 \left[y^i \frac{\partial \ln p(r;x)}{\partial x^i} \right] p(r;x) \frac{\partial \ln p(r;x)}{\partial x^l} dr. \quad (4-35)$$

Derivando a expressão anterior em relação a x^k e somando em y^k , teremos

$$L_{y^l x^k} y^k = \int_X 2 y^k y^i \frac{\partial^2 \ln p(r;x)}{\partial x^k \partial x^i} p(r;x) \frac{\partial \ln p(r;x)}{\partial x^l} dr +$$

$$\begin{aligned}
& + \int_X 2 \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right] \frac{\partial p}{\partial x^k} y^k \frac{\partial \ln p(r; x)}{\partial x^j} dr + \\
& + \int_X 2 \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right] p(r; x) y^k \frac{\partial^2 \ln p(r; x)}{\partial x^k \partial x^j} dr. \tag{4-36}
\end{aligned}$$

Derivando L em relação a x^j temos

$$L_{x^j} = \int_X 2 \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right] y^j \frac{\partial^2 \ln p(r; x)}{\partial x^j \partial x^i} p(r; x) dr + \int_X \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right]^2 \frac{\partial p(r; x)}{\partial x^j} dr. \tag{4-37}$$

Derivando N que é dado pela equação (3.32) em relação a y^j temos

$$\begin{aligned}
N_{y^j} &= \frac{1}{6} \int_X 3 \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right]^2 \frac{\partial y^j}{\partial y^j} \frac{\partial \ln p(r; x)}{\partial x^i} p(r; x) dr \\
&= \frac{1}{2} \int_X \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right]^2 \delta_{ij} \frac{\partial p(r; x)}{\partial x^i} dr \\
&= \frac{1}{2} \int_X \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right]^2 \frac{\partial p(r; x)}{\partial x^j} dr.
\end{aligned}$$

Agora

$$\begin{aligned}
L_{y^j x^k} y^k - L_{x^j} &= \int_X 2 y^k y^i \frac{\partial^2 \ln p(r; x)}{\partial x^k \partial x^i} p(r; x) \underbrace{\frac{\partial \ln p(r; x)}{\partial x^j}}_{\frac{\partial p(r; x)}{\partial x^j}} dr + \\
&+ \int_X 2 \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right] \underbrace{\frac{\partial p(r; x)}{\partial x^k}}_{p(r; x) \frac{\partial \ln p(r; x)}{\partial x^k}} y^k \underbrace{\frac{\partial \ln p(r; x)}{\partial x^j}}_{\frac{1}{p(r; x)} \frac{dp(r; x)}{dx^j}} dr - \\
&- 2 \int_X \frac{1}{2} \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right]^2 \frac{\partial p(r; x)}{\partial x^j} dr.
\end{aligned}$$

Teremos então

$$\begin{aligned}
L_{y^j x^k} y^k - L_{x^j} &= \int_X 2 y^k y^i \frac{\partial^2 \ln p(r; x)}{\partial x^k \partial x^i} \frac{\partial p(r; x)}{\partial x^j} + \\
&+ \int_X 2 \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right] \left[\frac{\partial \ln p(r; x)}{\partial x^k} y^k \right] \frac{dp(r; x)}{dx^j} dr - \\
&- \int_X \left[y^i \frac{\partial \ln p(r; x)}{\partial x^i} \right]^2 \frac{\partial p(r; x)}{\partial x^j} dr. \tag{4-38}
\end{aligned}$$

Logo

$$L_{y^j x^k} y^k - L_{x^j} = \int_X 2 y^k y^i \frac{\partial^2 \ln p(r; x)}{\partial x^k \partial x^i} \frac{\partial p(r; x)}{\partial x^j} dr +$$

$$\begin{aligned}
& + \int_X 2 \left[y^i \frac{\partial \ln p(r;x)}{\partial x^i} \right]^2 \frac{dp(r;x)}{dx^i} dr - \\
& - \int_X \left[y^i \frac{\partial \ln p(r;x)}{\partial x^i} \right]^2 \frac{\partial p(r;x)}{\partial x^i} dr.
\end{aligned} \tag{4-39}$$

teremos daí que

$$\begin{aligned}
L_{y^j, x^k} y^k - L_{x^j} &= \int_X 2 y^k y^j \frac{\partial^2 \ln p(r;x)}{\partial x^k \partial x^j} \frac{\partial p(r;x)}{\partial x^j} dr + \\
& + \int_X \left[y^j \frac{\partial \ln p(r;x)}{\partial x^j} \right]^2 \frac{dp(r;x)}{dx^j} dr.
\end{aligned}$$

Finalmente

$$L_{y^j, x^k} y^k - L_{x^j} = \int_X 2 y^k y^j \frac{\partial^2 \ln p(r;x)}{\partial x^k \partial x^j} \frac{\partial p(r;x)}{\partial x^j} + 2 N_{y^j}.$$

Voltando à expressão do α -spray, e fazendo $i = r$ com a mesma variação, teremos

$$\begin{aligned}
G_{\alpha, \rho}^i(x, y) &= \mathcal{G}^i(x, y) + \frac{\rho \alpha}{2} g^{ii}(x, y) N_{y^i}(x, y) \\
&= \frac{1}{4} g^{ii}(x) \left[2 \int_X y^k y^r \frac{\partial^2 \ln p(r;x)}{\partial x^k \partial x^r} \frac{\partial p(r;x)}{\partial x^i} dr \right] + \frac{g^{ii} 2 N_{y^i}}{4} + \frac{\rho \alpha}{2} g^{ii}(x) N_{y^i} \\
&= \frac{1}{2} g^{ii}(x) \left[\int_X y^k y^r \frac{\partial^2 \ln p(r;x)}{\partial x^k \partial x^r} \frac{\partial p(r;x)}{\partial x^i} dr \right] + \frac{1 + \rho \alpha}{4} g^{ii}(x) 2 N_{y^i} \\
&= \frac{1}{2} g^{ii}(x) \left[\int_X y^k y^r \frac{\partial^2 \ln p(r;x)}{\partial x^k \partial x^r} \frac{\partial p(r;x)}{\partial x^i} dr \right] + \frac{1 + \rho \alpha}{4} \int_X \left[y^r \frac{\partial \ln p(r;x)}{\partial x^r} \right]^2 \frac{\partial p(r;x)}{\partial x^i} dr
\end{aligned} \tag{4-40}$$

Teremos então

$$G_{\alpha, \rho}^i = \bar{G}^i + (\rho \alpha + 1) A^i \tag{4-41}$$

onde

$$\bar{G}^i = \frac{g^{ii}(x)}{2} \int_X \left[y^r y^j \frac{\partial^2 \ln p(r;x)}{\partial x^r \partial x^j} \right] \frac{\partial}{\partial x^i} p(r;x) dr \tag{4-42}$$

e

$$A^i = \frac{g^{ii}(x)}{4} \int_X \left[y^j \frac{\partial \ln p(r;x)}{\partial x^j} \right]^2 \frac{\partial}{\partial x^i} p(r;x) dr, \tag{4-43}$$

concluindo assim a demonstração do Teorema. \square

Observação 7 Podemos escrever L e N de formas diferentes. Fazamos $p(r;x) := p$.

observe que

$$\begin{aligned}
 L &= \int_X \left[\frac{\partial \ln p}{\partial x^i} y^i \right]^2 p dr \\
 &= \int_X y^j \frac{\partial}{\partial x^j} \left[\left(y^i \frac{\partial}{\partial x^i} \ln p \right) p dr \right] - \int_X \left[y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} \right] p dr \\
 &= \int_X y^i y^j \frac{\partial^2 p}{\partial x^i \partial x^j} dr - \int_X \left[y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} \right] p dr \\
 &= y^i y^j \frac{\partial^2}{\partial x^i \partial x^j} \int_X p dr - \int_X \left[y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} \right] p dr \\
 &= - \int_X \left[y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} \right] p dr.
 \end{aligned} \tag{4-44}$$

De maneira análoga, podemos escrever

$$\begin{aligned}
 6N &= \int_X \left[y^i \frac{\partial \ln p}{\partial x^i} \right]^3 p dr \\
 &= y^k \frac{\partial}{\partial x^k} \int_X \left[\frac{\partial \ln p}{\partial x^i} y^i \right]^2 p dr - 2 \int_X \left[y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} \right] \left[y^k \frac{\partial p}{\partial x^k} \right] dr,
 \end{aligned}$$

mas

$$y^k \frac{\partial}{\partial x^k} \int_X \left[y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} p \right] dr = \int_X y^i y^j y^k \frac{\partial^3 \ln p}{\partial x^i \partial x^j \partial x^k} p dr + \int_X y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} y^k \frac{\partial p}{\partial x^k} dr. \tag{4-45}$$

Utilizando esse fato e (4-16) temos

$$\begin{aligned}
 6N &= y^k \frac{\partial}{\partial x^k} \int_X \left[\frac{\partial \ln p}{\partial x^i} y^i \right] \left[y^j \frac{\partial p}{\partial x^j} \right] dr - 2 y^k \frac{\partial}{\partial x^k} \int_X \left[y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} p \right] dr + \\
 &\quad + 2 \int_X y^i y^j y^k \frac{\partial^3 \ln p}{\partial x^i \partial x^j \partial x^k} p dr.
 \end{aligned}$$

Utilizando (4-44), temos

$$6N = -3 y^k \frac{\partial}{\partial x^k} \int_X \left[y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} p \right] dr + 2 \int_X y^i y^j y^k \frac{\partial^3 \ln p}{\partial x^i \partial x^j \partial x^k} p dr. \tag{4-46}$$

Essas expressões serão utilizadas na próxima seção.

4.3 Família exponencial de distribuições de probabilidade

Definição 4.2 Uma variedade M em \mathcal{P} é chamada de *família exponencial* ou *variedade exponencial* se é coberta por injeções

$$\bar{\omega} : \Omega \subset \mathbb{R}^n \rightarrow M \subset \mathcal{P},$$

tal que, $p := \bar{\omega}(x) \in \mathcal{P}$ pode ser expressa por

$$p(r; x) = e^{x^i f_i(r) + k(r) - \psi(x)}. \quad (4-47)$$

Nessa seção, defina $p := p(r, x)$. Observe que $\int_X \frac{\partial p}{\partial x^i} dr = 0$. Por outro lado,

$$\begin{aligned} \int_X \frac{\partial p}{\partial x^i} dr &= \int_X \left(f_i(r) - \frac{\partial \psi(x)}{\partial x^i} \right) p dr \\ &= \int_X f_i(r) p dr - \frac{\partial \psi(x)}{\partial x^i} \int_X p dr. \end{aligned}$$

portanto

$$\frac{\partial \psi(x)}{\partial x^i} = \int_X f_i(r) p dr. \quad (4-48)$$

Exemplo 4.3

A divergência de *Kullback-Leibler* D_{KL} em M é a f -divergência com $f(t) = -\ln t$. Sendo assim,

$$\begin{aligned} D_{KL}(p, p') &= \int_X p(r) \ln \left(\frac{p(r)}{p'(r)} \right) dr \\ &= \int_X p(r) \ln \left(\frac{e^{x^i f_i(r) + k(r) - \psi(x)}}{e^{x'^i f_i(r) + k(r) - \psi(x')}} \right) dr \\ &= \int_X p [x^i f_i - k - \psi(x) - x'^i f_i + k + \psi(x')] dr \\ &= \int_X p [(x^i - x'^i) f_i + \psi(x') - \psi(x)] dr \\ &= \psi(x') \int_X p dr - \psi(x) \int_X p dr - (x' - x)^i \int_X p f_i(r) dr \\ &= \psi(x') - \psi(x) - (x' - x)^i \frac{\partial \psi(x)}{\partial x^i}. \end{aligned} \quad (4-49)$$

Proposição 4.4 *Seja M uma família exponencial de distribuições de probabilidade como em (3.44). A estrutura de informação induzida de D_f é dada por $(L_f, H_f) = (L, \rho N)$ com $\rho = 3 + 2f'''(1)$ e*

$$L = \frac{\partial^2 \Psi(x)}{\partial x^i \partial x^j} y^i y^j, \quad (4-50)$$

$$N = \frac{1}{6} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^k} y^i y^j y^k. \quad (4-51)$$

Prova.

Notemos que $\ln p(r, x) = x^i f_i(r) + k(r) - \psi(x)$. sabemos que

$$L = - \int_{\mathcal{X}} y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} p dr.$$

Então

$$\begin{aligned} \frac{\partial^2 \ln p}{\partial x^i \partial x^j} &= \frac{\partial}{\partial x^i} \left[\frac{\partial}{\partial x^j} (x^i f_i(r) + k(r) - \psi(x)) \right] \\ &= \frac{\partial}{\partial x^i} \left(\delta_{ij} f_i(r) - \frac{\partial \psi(x)}{\partial x^j} \right) \\ &= - \frac{\partial^2 \psi(x)}{\partial x^i \partial x^j}. \end{aligned}$$

Logo

$$\begin{aligned} L &= - \int_{\mathcal{X}} y^i y^j \frac{\partial^2 \ln p}{\partial x^i \partial x^j} p dr \\ &= \frac{\partial^2 \psi(x)}{\partial x^i \partial x^j} y^i y^j \int_{\mathcal{X}} p dr \\ &= \frac{\partial^2 \psi(x)}{\partial x^i \partial x^j} y^i y^j. \end{aligned}$$

Derivemos L em relação a x^k , assim teremos

$$L_{x^k} = \frac{\partial^3 \psi(x)}{\partial x^i \partial x^j \partial x^k} y^i y^j. \quad (4-52)$$

Derivando L_{x^k} em relação a y^l e somando em y^k teremos

$$\begin{aligned} L_{x^k, y^l, y^k} &= \frac{\partial^3 \psi(x)}{\partial x^k \partial x^l \partial x^j} y^j y^k + \frac{\partial^3 \psi(x)}{\partial x^k \partial x^i \partial x^j} y^i y^k \\ &= 2 \frac{\partial^3 \psi(x)}{\partial x^i \partial x^j \partial x^k} y^i y^j. \end{aligned} \quad (4-53)$$

Logo, a expressão dos coeficientes do spray de L é dada por

$$G^i = \frac{1}{4} g^{ik}(x) \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^k} y^j y^i. \quad (4-54)$$

Como N pode ser escrito como em (4.3), teremos então

$$\begin{aligned} N &= \frac{1}{3} \int_X y^i y^j y^k \frac{\partial^3 \ln p}{\partial x^i \partial x^j \partial x^k} p dr - \frac{1}{2} y^k \frac{\partial}{\partial x^k} \int_X y^i y^j \frac{\partial^3 \ln p}{\partial x^i \partial x^j} p dr \\ &= \frac{1}{3} \int_X y^i y^j y^k \frac{\partial}{\partial x^i} \left\{ \frac{\partial}{\partial x^j} \left[\frac{\partial}{\partial x^k} (x^j f_i(r) + k(r) - \Psi(x)) \right] \right\} p dr + \frac{1}{2} y^k \frac{\partial}{\partial x^k} \left(\frac{\partial^2 \Psi(x)}{\partial x^i \partial x^j} y^i y^j \right) \\ &= \frac{1}{3} y^i y^j y^k \frac{\partial}{\partial x^i} \left[\frac{\partial}{\partial x^j} \left(f_k - \frac{\partial \Psi(x)}{\partial x^k} \right) \right] + \frac{1}{2} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^k} y^i y^j y^k \\ &= \frac{1}{3} \int_X y^i y^j y^k \frac{\partial}{\partial x^i} \left(-\frac{\partial^2 \Psi(x)}{\partial x^j \partial x^k} \right) + \frac{1}{2} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^k} y^i y^j y^k \\ &= -\frac{1}{3} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^k} y^i y^j y^k + \frac{1}{2} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^k} y^i y^j y^k \\ &= \frac{1}{6} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^k} y^i y^j y^k. \end{aligned} \quad (4-55)$$

□

Corolário 4.5 *Seja M uma família exponencial de distribuições de probabilidade na forma (4-47). Seja (L_f, H_f) a estrutura de informação induzida pela f -divergência. Quando $\rho\alpha = -1$, temos que (L_f, H_f) é α -flat, isto é, o α -spray de (L_f, H_f) é flat.*

Prova.

O α -spray é dado por

$$G_{\alpha, \rho}^i(x, y) = G^i(x, y) + \frac{\alpha \rho}{2} g^{il}(x, y) N_{y^l}(x, y) \quad (4-56)$$

Derivando N dado por em relação a y^l vamos obter

$$\begin{aligned} N_{y^l} &= \frac{1}{6} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^k} \left[\delta_{il} y^j y^k + \delta_{jl} y^i y^k + \delta_{kl} y^i y^j \right] \\ &= \frac{1}{6} [3y^l y^l] \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^l} \\ &= \frac{1}{2} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^l} y^i y^j \end{aligned}$$

logo

$$\begin{aligned} G_{\alpha,p}^i(x,y) &= \mathcal{G}^i = \frac{1}{4} g^{ij}(x) \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^l} y^j y^l + \frac{\alpha p}{2} g^{il} \frac{1}{2} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^l} y^j y^l \\ &= \frac{\rho \alpha + 1}{4} g^{il} \frac{1}{2} \frac{\partial^3 \Psi(x)}{\partial x^i \partial x^j \partial x^l} y^j y^l. \end{aligned}$$

logo, para $\rho \alpha = -1$, temos que $G_{\alpha,p}^i = 0$, ou seja, o α -spray da estrutura é flat. \square

No exemplo a seguir, iremos tomar uma família M de distribuições normais, parametrizada no aberto

$$U = \{(\mu, \sigma) : \mu \in \mathbb{R}, \sigma \in \mathbb{R}_+\}.$$

Verificaremos que M é um espaço de curvatura gaussiana negativa, determinando a matriz de sua métrica, utilizando os fatos provados anteriormente.

Exemplo 4.6 Considere agora uma família M de distribuições normais gaussianas com média μ e variância σ^2 :

$$p(r; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} e^{\left[-\frac{1}{2} \frac{(r-\mu)^2}{\sigma^2}\right]}.$$

Vamos reparametrizar M por

$$p(r; x) = e^{x^1 f_1(r) + x^2 f_2(r) + k(r) - \Psi(x)},$$

onde

$$x^1 = \frac{\mu}{\sigma^2}, \quad x^2 = \frac{1}{2\sigma^2}, \quad f_1(r) = r, \quad f_2(r) = -r^2 \quad e \quad k(r) = 0.$$

Façamos

$$e^{x^1 f_1(r) + x^2 f_2(r) + k(r) - \Psi(x)} = \frac{1}{\sqrt{2\pi\sigma}} e^{\left[-\frac{(r-\mu)^2}{\sigma^2}\right]}.$$

Teremos

$$e^{\frac{\mu}{\sigma^2} r + \frac{1}{2\sigma^2} (-r^2) - \Psi(x)} = \frac{1}{\sqrt{2\pi\sigma}} e^{\left[-\frac{(r-\mu)^2}{\sigma^2}\right]}.$$

Dessa igualdade, obteremos

$$\Psi(x) = \frac{\mu^2}{2\sigma^2} + \log(\sqrt{2\pi\sigma}) = \frac{(x^1)^2}{4x^2} \ln \sqrt{\frac{\pi}{x^2}}. \quad (4-57)$$

Assim M é uma variedade exponencial em \mathcal{P} e

$$L = \frac{\partial^2 \Psi(x)}{\partial x^i \partial x^j} y^i y^j,$$

onde

$$g_{ij}(x) = \frac{\partial^2 \psi(x)}{\partial x^i \partial x^j}.$$

Para a variedade M , obtemos a matriz da métrica:

$$[g_{ij}(x)] = \begin{bmatrix} \frac{1}{2x^2} & -\frac{x^1}{2(x^2)^2} \\ -\frac{x^1}{2(x^2)^2} & \frac{(x^1)^2}{2(x^2)^3} + \frac{1}{2x^2} \end{bmatrix},$$

calculando a curvatura Gaussiana dessa métrica, obteremos $K = -\frac{1}{2}$.

No próximo exemplo, iremos tomar S como uma família de distribuições gamma, e verificaremos que S é um espaço de curvatura gaussiana negativa, determinando um intervalo para sua curvatura gaussiana.

Exemplo 4.7 Seja S a família das distribuições gamma com espaço de eventos $\Omega = \mathbb{R}^+$ e parâmetros $\tau, \nu \in \mathbb{R}^+$ definida por

$$p(r; \tau, \nu) = \left(\frac{\nu}{\tau}\right)^\nu \frac{r^{\nu-1}}{\Gamma(\nu)} e^{-\frac{r\nu}{\tau}}, \quad (4-58)$$

onde τ e ν variam no conjunto

$$U = \left\{ (\mu, \nu) : \mu = \frac{\nu}{\tau} \in \mathbb{R}, \nu \in \mathbb{R}^+ \right\},$$

e Γ é a função-gamma definida por

$$\Gamma(\nu) = \int_0^\infty s^{\nu-1} e^{-s} ds. \quad (4-59)$$

Primeiramente, vamos reparametrizar M colocando $x^1 = \mu$, $x^2 = \nu$, $f_1 = -r$, $f_2 = \ln r$ e $k(r) = \ln r$.

Utilizando a equação (4-47), temos

$$p(r; \mu, \nu) = e^{-\mu r + \nu \ln r - \ln r - \psi} = \mu^\nu \frac{r^{\nu-1}}{\Gamma(\nu)} e^{-r\mu}.$$

Aplicando o logaritmo em ambos os lados dessa equação, teremos

$$-\mu r + \nu \ln r - \ln r - \psi = \nu \ln(\mu) + (\nu - 1) \ln r - \ln \Gamma(\nu) - r\mu,$$

assim

$$\psi = \ln \Gamma(\nu) - \nu \ln \mu.$$

Logo S é uma variedade exponencial em \mathcal{P} e

$$g_{11} = \frac{\partial^2 \psi}{\partial \mu \partial \mu} = \left(\frac{-v}{\mu} \right)_{\mu} = \frac{v}{\mu^2},$$

$$g_{12} = g_{21} = \frac{\partial^2 \psi}{\partial \mu \partial v} = \left(\frac{-v}{\mu} \right)_{v} = -\frac{1}{\mu},$$

$$g_{22} = \frac{\partial^2 \psi}{\partial v \partial v} = \left(\frac{\Gamma'(v)}{\Gamma(v)} \right)_{v} = \kappa'(v),$$

onde $\kappa(v) = \frac{\Gamma'(v)}{\Gamma(v)}$.

Assim temos a seguinte métrica:

$$ds^2 = \frac{v}{\mu^2} d\mu^2 - \frac{2}{\mu} d\mu dv + \kappa'(v) dv^2. \quad (4-60)$$

Fazendo agora a mudança de coordenadas

$\beta = \frac{v}{\mu}$ e $v = v$, teremos

$$d\mu = \frac{1}{\beta} dv - \frac{v}{\beta^2} d\beta.$$

Utilizando esse fato, a métrica assumirá a forma:

$$ds^2 = \frac{\beta^2}{v} \left(\frac{1}{\beta} dv - \frac{v}{\beta^2} d\beta \right) - \frac{2\beta}{v} \left(\frac{1}{\beta} dv - \frac{v}{\beta^2} d\beta \right) dv + \kappa'(v) dv^2,$$

então,

$$ds^2 = \frac{v}{\beta^2} d\beta^2 + \left(\kappa'(v) - \frac{1}{v} \right) dv^2. \quad (4-61)$$

Com essa métrica, teremos a seguinte matriz da métrica:

$$[g_{ij}] = \begin{bmatrix} \frac{v}{\beta^2} & 0 \\ 0 & \left(\kappa'(v) - \frac{1}{v} \right) \end{bmatrix}.$$

Voltando para o sistema de coordenadas anterior, temos

$$[g_{ij}] = \begin{bmatrix} \frac{v}{\mu^2} & 0 \\ 0 & \left(\kappa'(v) - \frac{1}{v} \right) \end{bmatrix}.$$

Voltando ainda para o sistema de coordenadas (x^i) , teremos finalmente

$$[g_{ij}(x)] = \begin{bmatrix} x^2 x^1 & 0 \\ 0 & \left(\kappa'(x^2) - \frac{1}{x^2} \right) \end{bmatrix}.$$

Calculando-se os Símbolos de Christoffel para a métrica dada pela equação (4-60), teremos

$$\Gamma_{11}^1 = \frac{1 - 2v\kappa'(v)}{2\mu(-1 + v\kappa'(v))},$$

$$\Gamma_{12}^1 = \frac{\kappa'(v)}{2(-1 + v\kappa'(v))},$$

$$\Gamma_{22}^1 = \frac{\mu\kappa''(v)}{2(-1 + v\kappa'(v))},$$

$$\Gamma_{11}^2 = \frac{v}{2\mu^2(1 - v\kappa'(v))},$$

$$\Gamma_{12}^2 = \frac{1}{2\mu(-1 + v\kappa'(v))}$$

$$\Gamma_{22}^2 = \frac{v\kappa''(v)}{2(-1 + v\kappa'(v))}.$$

Utilizando a Equação de Gauss (Ver em [2]) e fazendo as simplificações necessárias, veremos que a curvatura gaussiana dessa variedade será dada por

$$K = \frac{\kappa'(v) + v\kappa''(v)}{4(v\kappa'(v) - 1)^2}.$$

Segundo [8], a função K satisfaz

$$-\frac{1}{2} < K < -\frac{1}{4}.$$

4.4 Dualidade das f -divergências em espaços de probabilidade

Essa seção tem por objetivo, aplicar o Teorema 3.24 para as f -divergências, ou seja provar que a estrutura de informação induzida pela f -divergência dada por (L_f, H_f) é α -flat se, e somente se, $(L_{f^*}, H_{f^*}) = (L_f(x, -y), H_f(x, -y))$ é α -flat.

Definição 4.8 *Seja $(\mathcal{P}, \mathcal{D})$ um espaço de divergência. Definimos a divergência dual D^* em \mathcal{P} por*

$$D^*(p, q) = D(q, p), \forall p, q \in \mathcal{P}.$$

Dada uma função convexa $f : (0, \infty) \rightarrow \mathbb{R}$ com $f(1) = 0$ e $f''(1) = 1$, seja

$$f^*(t) := tf\left(\frac{1}{t}\right), t > 0.$$

Então

$$f^*(1) = 1f(1) = 0.$$

Calculando suas derivadas primeira, segunda e terceira em relação a t , temos:

$$f^{*'}(t) = f\left(\frac{1}{t}\right) - \frac{1}{t}f'\left(\frac{1}{t}\right);$$

$$f^{**}(t) = \frac{1}{t^3}f''\left(\frac{1}{t}\right);$$

$$f^{***}(t) = -\frac{3}{t^4}f''\left(\frac{1}{t}\right) - \frac{1}{t^5}f'''(1).$$

Fazendo $t = 1$, então:

$$f^{**}(1) = 1$$

e

$$f^{***}(1) = -3f''(1) - f'''(1) = -3 - f'''(1).$$

Sendo

$$\rho = 3 + 2f'''(1),$$

como no Teorema 4.1, faremos

$$\rho^* = 3 + 2f^{***}(1) = 3 + 2(-3 - f'''(1)) = -3 - 2f'''(1) = -\rho.$$

Logo

$$\rho^* + \rho = 0.$$

Olhando para a f -divergência em espaços de probabilidade, temos

$$(D_f)^*(p, q) = D_f(q, p) = \int_X q(r) f\left(\frac{p(r)}{q(r)}\right) dr$$

e

$$\begin{aligned} D_{f^*}(p, q) &= \int_X p(r) f^*\left(\frac{q(r)}{p(r)}\right) dr \\ &= \int_X q(r) \frac{p(r)}{q(r)} f^*\left(\frac{1}{\frac{p(r)}{q(r)}}\right) dr \\ &= \int_X q(r) f\left(\frac{p(r)}{q(r)}\right) \\ &= D_f(q, p). \end{aligned}$$

Então $(D_f)^* = D_{f^*}$ induz, segundo o Teorema 4.1, uma estrutura de informação

$$(L_{f^*}, H_{f^*}) = (L, \rho^* N) = (L, -\rho N),$$

onde

$$L_{f^*}(x, y) = L_f(x, -y) = \int_{\mathcal{X}} \left[y^i \frac{\partial}{\partial x^i} \ln p \right]^2 p dr,$$

e

$$H_{f^*}(x, y) = H_f(x, -y) = -\frac{\rho}{6} \int_{\mathcal{X}} \left[y^i \frac{\partial}{\partial x^i} \ln p \right]^3 p dr.$$

A estrutura de informação induzida de $(D_f)^*$ é a dual de D_f . Nesse sentido D_f é chamada de *dualística*.

Proposição 4.9 *A estrutura de informação (L_f, H_f) é α -flat se, e somente se, $(L_{f^*}, H_{f^*}) = (L_f(x, -y), H_f(x, -y))$ é α -flat.*

Prova. Sabemos que a estrutura de informação dual de (L_f, H_f) é $(L_{f^*}, H_{f^*}) = (L_f(x, -y), H_f(x, -y))$. Utilizando-se o Teorema 3.24, segue diretamente a demonstração da Proposição. \square

Referências Bibliográficas

- [1] Bao, D., Chern, S.S., Shen, Z., *An Introduction to Riemann-Finsler Geometry*, GTM **200** Springer-Verlag,(2000).
- [2] do Carmo, M.P., *Differential geometry of curves and surfaces*, Prentice Hall.
- [3] do Carmo, M.P., *Geometria Riemanniana*, Projeto Euclides.
- [4] Shen, Z., *Riemann-Finsler Geometry with applications to Information Geometry*, Chin. Ann. Math. **27B(1)** (2006), 73-94.
- [5] Atkinson, C. and Mitchell,A.F.S., *Rao's Distance Measure*, Sankhya: The Indian Journal of Statistics ,**43A(3)** ,(1981), pp. 345-365.
- [6] Amari, S.-I., Nagaoka, H., *Methods in Information Geometry*, Oxford University Press and Ameri. Math. Soc., (2000),
- [7] Amari, S.-I., *Differential Geometrical Methods in Statistics*, Springer Lecture notes in Statistics , **20**, Springer , (2002).
- [8] Dodson, C.T.J., Matsusoe, H., *An affine embedding of the Gamma Manifold*, Appl. Sci. (eletronic), **5(1)**, (2003), 7-12.
- [9] Shen, Z., *Lectures on Finsler Geometry*, Department of Mathematical Sciences.
- [10] Lima, E.L., *Análise no Espaço \mathbb{R}^n* , Projeto Euclides.
- [11] Lima, E.L., *Variedades Diferenciáveis*, Projeto Euclides.
- [12] Corruera ,J.M., Giummolè ,F., *A characterization of Monotone and Regular Divergences*, Ann. Inst. Statist. Math. **50(3)**,(1998) , 433-450.
- [13] Nagoka , H., Amari , S.-I., *Differential geometry of smooth families of probability distributions*, Univ. Tokyo , Japan , METR, (1982) , 82-87.
- [14] Shen, Z., *Differential Geometry of Spray and Finsler Spaces* , Kluwer Academic Publishers, Dordrecht, (2001).

- [15] James, B., *Probabilidade: um curso em nível intermediário*, segunda edição, projeto euclides, IMPA, Brasil, (1996).
- [16] Fernandez, P.J. *Medida e Integração*, segunda edição, projeto euclides, IMPA, Brasil, (2002).
- [17] Hoel, P.G., Port, S.C., Stone, C.J., *Introdução à teoria da probabilidade*, UC, Los Angeles, tradução : Chiyoshi, F.Y., editora intercência, Brasil, (1971).
- [18] Chern, S.S. *Finsler Geometry is just Riemannian geometry without the quadratic restriction*, Notices AMS. **43**, (1996), 9, 959-963.

Notações

Seguem aqui uma lista de notações utilizadas durante o texto.

1. \mathbb{E} significa *Esperança matemática de uma variável aleatória*.
2. G^i significa *coeficientes do spray G* .
3. g_{ij} significa *Coefficientes da matriz da métrica em questão*.
4. g^{ij} significa *Coefficientes da matriz inversa da matriz da métrica em questão*.
5. δ_{ij} significa *delta de Kronecker*.
6. $L(x,y)$ significa *Expressão da métrica de Finsler*.
7. $L(c)$ significa *Comprimento da curva c* .
8. $H(x,y)$ significa *Expressão da H-função*.
9. γ_{ij}^k significa *símbolos de Christoffel para uma métrica Riemanniana*.
10. Γ_{ij}^k significa *símbolos de Christoffel para um spray*.
11. \mathbb{P} significa *medida de probabilidade*.
12. Ω significa *espaço amostral*.
13. $\Gamma(\alpha)$ significa *Função gamma com parâmetro α* .

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)