

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**De modelos estatísticos a α -conexões: um panorama geral
da geometria da informação**

Gabriel dos Reis Trindade

Dissertação de Mestrado do Programa de Pós-Graduação em
Matemática (PPG-Mat)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Gabriel dos Reis Trindade

De modelos estatísticos a α -conexões: um panorama geral da geometria da informação

Dissertação apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Mestre em Ciências – Matemática. *VERSÃO REVISADA*

Área de Concentração: Matemática

Orientador: Prof. Dr. Carlos Henrique Grossi Ferreira

USP – São Carlos
Setembro de 2023

Ficha catalográfica elaborada pela Biblioteca Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

T833m Trindade, Gabriel dos Reis.
De modelos a α -conexões: um panorama geral da geometria da
informação / Gabriel dos Reis Trindade; orientador Carlos Henrique Grossi
Ferreira. -- São Carlos, 2023
137 p.

Dissertação (Mestrado – Programa de Pós-Graduação em Matemática). –
Instituto de Ciências Matemáticas e de Computação, Universidade de São
Paulo, 2023.

1. Geometria da informação. 2. Métrica de Fisher. 3. Tensor de Amari-
Chentsov. 4. Modelos estatísticos. 5. Estatística. I. Ferreira, Carlos Henrique
Grossi, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Gabriel dos Reis Trindade

From statistical models to α -connections: an overview of
information geometry

Master dissertation submitted to the Instituto de
Ciências Matemáticas e de Computação – ICMC-
USP, in partial fulfillment of the requirements for the
degree of the Master Program in Mathematics. *FINAL
VERSION*

Concentration Area: Mathematics

Advisor: Prof. Dr. Carlos Henrique Grossi Ferreira

USP – São Carlos
September 2023

*Este trabalho é dedicado a todos que lutam para que
aqueles amados por eles possam estar presentes na universidade.
Em especial, aos meus pais: Maria Aparecida Salomoni Trindade e Valdeci Trindade.*

AGRADECIMENTOS

Em primeiro lugar, agradeço aos meus pais, Maria Aparecida e Valdeci, pelos seus esforços veementes para garantir que meus sonhos pudessem vir a se realizar, e por toda confiança que tiveram de que eu faria esses esforços serem recompensados. Junto a eles, estão meus familiares próximos, para com os quais eu tenho enorme gratidão. Entre eles, destaco minhas irmãs: Lediane e Roselaine; e meu cunhado, Cristiano, por terem me acompanhado durante todas as etapas do mestrado. Destaco, também, meus sobrinhos: Giovana, Guilherme e Vitor, com os quais dividi sorrisos restauradores durante os momentos mais desafiadores dentro da universidade. Agradeço, igualmente, à segunda família que me abraçou, sendo encabeçada pela minha Tatá e meu Vô. E, mantendo-me na esfera da família, tanto daquela do presente quanto do futuro, sou grato ao Daniel, que cotidianamente esteve ao meu lado, ajudando-me diretamente com problemas relativos à matemática e, ao mesmo tempo, dando-me incentivo constante para eu ser melhor.

Para além da minha família, expresso minha completa gratidão aos meus amigos, sem os quais, nem este trabalho, nem o Gabriel que está escrevendo, seriam os mesmos. Chamo atenção à Juliana e à Nayla, que me são tão próximas e amadas que se tornaram um pedaço de mim mesmo; ao Gustavo Habermann, cujas risadas que compartilhamos sempre vieram embebidas de uma boa dose de matemática; ao Matheus e à Clara, que há anos permanecem ao meu lado; à Fernanda, que sempre esteve com as mãos estendidas para me ajudar; ao Victor e ao Thiago, que foram dois dos melhores presentes que o ICMC poderia me dar; ao Saulo, que constantemente se prontificou a me ajudar na busca por saídas para diversas dificuldades; ao Hugo Botós e ao Gabriel Magno, que estiveram dispostos a me auxiliar com as dúvidas que eu tinha, sendo eles primordiais para a realização desta dissertação; aos amigos geniais de laboratório: Aires, Alan, Arnando, Gustavo, Julia, Pedro e Sérgio; e a todo o grupo do 61 - 16: Leonardo, Felipe, Felipe Picoli, Giuliano, Guilherme Fiusa, Gustavo e Mateus Otto, que tenho muito orgulho de chamar de amigos e com quem compartilhei alguns dos melhores momentos da minha vida. Expresso, também, minha gratidão aos docentes que, em maior ou menor quantidade, deixaram um pedaço de si neste trabalho. Enfatizo os professores Diogo Pinto, Leonardo Maia e Carlos Grossi, esse último que, com imenso orgulho, pude chamar de orientador. Sem ele, este trabalho não existiria, assim como, provavelmente, essa minha vívida curiosidade matemática, sendo, ele, uma inspiração para qual pesquisador e docente quero ser. Agradeço, também, à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), processo 2021/12253-7, pelo fomento. Finalmente, o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

*“nunca sei ao certo
se sou um menino de dúvidas
ou homem de fé*

*certezas o vento leva
só dúvidas ficam de pé.”
(Paulo Leminski)*

RESUMO

TRINDADE, G. R. **De modelos estatísticos a α -conexões: um panorama geral da geometria da informação**. 2023. 137 p. Dissertação (Mestrado em Ciências – Matemática) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Esse trabalho é uma apresentação sobre a geometria da informação, organizando-se como um compilado de conceitos e resultados fundamentais da área, bem como aplicações em teoria de informação quântica. Dito isso, é possível apresentar a geometria da informação entendendo-a como a área que se utiliza de ferramentas da geometria diferencial, em especial da geometria riemanniana, para resolver problemas advindos da estatística. Possuindo um caráter interdisciplinar que transpassa seu desenvolvimento histórico, ela interpreta modelos estatísticos como variedades diferenciáveis, munindo-os de uma métrica riemanniana e de um campo tensorial 3-covariante, chamados, respectivamente de métrica de Fisher e de tensor de Amari-Chentsov. No contexto de geometria da informação finita, eles são os únicos campos tensoriais covariantes de *rank* 3 invariantes por morfismos de Markov induzidos por núcleos de Markov congruentes. Além disso, dentre as famílias de distribuições de probabilidade, aquela constituída pelas distribuições gaussianas apresenta uma métrica de Fisher com um formato conhecido e, conseqüentemente, uma geometria familiar, sendo essa a geometria hiperbólica bidimensional. Ademais, os modelos estatísticos podem ser munidos de uma família uniparamétrica de conexões, chamadas de α -conexões, dentre as quais se destacam, especialmente no contexto finito, a conexão mistura e a exponencial. Já partindo de uma variedade munida de uma métrica riemanniana g e de um tensor 3-simétrico T , é possível induzir um par de conexões lineares livres de torção nela e que se relacionam através de um enfraquecimento da noção de compatibilidade com a métrica, chamadas de conexões duais uma a outra em relação a g . Elas, por sua vez, junto à métrica riemanniana, induzem um tensor 3-simétrico na variedade. Então, por meio do estudo dessas conexões, tem-se que a geometria que emerge da combinação de uma métrica riemanniana g com duas conexões planas ∇ e ∇^* duais uma a outra em relação a g é equivalente, ao menos localmente, a uma única função convexa, onde essa convexidade é considerada em relação a um sistema de coordenadas afim para uma das conexões duais. Outrossim, podendo ser aplicada à teoria de informação quântica a fim de se obter limites de velocidade quânticos geométricos, a geometria da informação conduz a generalizações do princípio de incerteza para a energia e o tempo em sistemas quânticos, em que o análogo quântico da métrica de Fisher clássica produz tais limites.

Palavras-chave: Geometria da informação, Métrica de Fisher, Tensor de Amari-Chentsov, Modelos estatísticos, Estatística.

ABSTRACT

TRINDADE, G. R. **From statistical models to α -connections: an overview of information geometry**. 2023. 137 p. Dissertação (Mestrado em Ciências – Matemática) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

This paper is a presentation on information geometry, organized as a compilation of fundamental concepts and results of the area, as well as applications in quantum information theory. Nevertheless, it is possible to present the information geometry by regarding it as the area that uses tools from differential geometry, especially from Riemannian geometry, to solve problems arising from statistics. Possessing an interdisciplinary character that permeates its historical development, it interprets statistical models as differentiable manifolds, providing them with a Riemannian metric and a 3-covariant tensor field, called, respectively, Fisher metric and Amari-Chentsov tensor. In the context of finite information geometry, they are the only covariant tensor fields of *rank* 3 invariant by Markov morphisms induced by congruent Markov kernels. Furthermore, among the families of probability distributions, the one formed by the Gaussian distributions has Fisher metric with a known structure and, consequently, a familiar geometry, which is the two-dimensional hyperbolic geometry. In addition, statistical models can be equipped with a uniparametric family of connections, called α -connections, among which, especially in the finite context, the mixed connection and the exponential connection stand out. Moreover, starting with a manifold equipped with a Riemannian metric g and a 3-symmetric tensor T , it is possible to induce a pair of torsion-free linear connections on it and that are related by a weakening of the notion of compatibility with the metric; that are called dual connections to each other with respect to g . They, in turn, together with the Riemannian metric, induce a 3-symmetric tensor on the manifold. Then, through the study of these connections, the geometry that emerges from the combination of a Riemannian metric g with two flat connections ∇ and ∇^* dual to each other with respect to g is equivalent, at least locally, to a single convex function, where this convexity is considered with respect to an affine coordinate system for one of the dual connections. Furthermore, being able to be applied to quantum information theory in order to obtain geometric quantum speed limits, information geometry leads to generalizations of the uncertainty principle for energy and time in quantum systems, where the quantum analog of the metric of classical Fisher produces such limits.

Keywords: Information geometry, Fisher metric, Amari-Chentsov tensor, Statistical models, Statistics.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação gráfica de um modelo estatístico gaussiano. Cada elemento do semiplano euclidiano superior pode ser observado como uma distribuição gaussiana	36
Figura 2 – Representação gráfica de diferentes medidas de dissimilaridade entre duas distribuições p_{ξ_1} e p_{ξ_2} de um modelo estatístico S do ponto de vista da α -geometria. Supondo suas existências, $\gamma^{(0)}$, $\gamma^{(-\alpha)}$ e $\gamma^{(\alpha)}$ representam, respectivamente, geodésicas relativas a $\nabla^{(0)}$, $\nabla^{(-\alpha)}$ e $\nabla^{(\alpha)}$ ligando essas distribuições	46
Figura 3 – Representação gráfica da semelhança entre \mathbb{H}^2 e \mathbb{E} no caso da família de distribuições gaussianas. Em \mathbb{H}^2 , estão ilustradas as curvas γ_1 e γ_2 , que contemplam o formato das geodésicas de $(\mathbb{H}^2, g_{\mathbb{H}^2})$	48
Figura 4 – Representação gráfica de $\mathcal{P}_+(\mathcal{X})$ para $n = 2$. É possível observar o aspecto de 2-simplexo de $\mathcal{P}_+(\mathcal{X})$	55
Figura 5 – Representação gráfica do difeomorfismo $\pi^{1/2}$ entre $\mathcal{P}_+(\mathcal{X})$ e $S_{2,+}(\mathcal{X})$	58
Figura 6 – Representação gráfica das relações entre \mathcal{X} , \mathcal{X}' , $\mathcal{P}(\mathcal{X})$ e $\mathcal{P}(\mathcal{X}')$ através de núcleos de Markov e morfismos de Markov	59
Figura 7 – Representação gráfica das relações entre \mathcal{X} , \mathcal{X}' , $\mathcal{P}(\mathcal{X})$ e $\mathcal{P}(\mathcal{X}')$ através de núcleos de Markov congruentes com respeito à estatística $\kappa: \mathcal{X} \rightarrow \mathcal{X}'$	64
Figura 8 – Exemplo de m - e e -geodésica em $\mathcal{M}_+(\mathcal{X})$ no caso de $\dim \mathcal{M}_+(\mathcal{X}) = 2$. Como $\tilde{\nabla}^{(m)}$ possui símbolos de Christoffel nulos, suas geodésicas são retas em $\mathcal{M}_+(\mathcal{X})$, enquanto as de $\tilde{\nabla}^{(e)}$ são exponenciais	70
Figura 9 – Exemplo de m - e e -geodésica em $\mathcal{P}_+(\mathcal{X})$ no caso de $\dim \mathcal{P}_+(\mathcal{X}) = 2$	75
Figura 10 – Representação gráfica de α -geodésicas com $\dim \mathcal{M}_+(\mathcal{X}) = 2$ e $\alpha = -1, -\frac{1}{2}, 0$ e 1 . Para que se possa identificar as proporção, foram tomados $\mu = (1, 2)$ e $a = (-4, 5)$	84
Figura 11 – Representação gráfica de uma subvariedade autoparalela S de M . Denotando por $T_p S^\perp$ o complemento ortogonal de $T_p S$ em $T_p M$, como $\nabla_A B \in \mathfrak{X}(S)$ para todo $A, B \in S$, a projeção de $\nabla_A B(p)$ em $T_p S^\perp$ é nula para todo $p \in S$, o que motiva a nomenclatura para S	97
Figura 12 – Representação gráfica de P como argumento de minimização da intersecção de H_Q com o eixo z , para Q ponto do gráfico da função estritamente convexa F e H_Q hiperplano de inclinação $grad_{x(P)}(F)$ passando por Q	100

Figura 13 – Representação gráfica de uma variação analítica na evolução dinâmica de ρ_0 . As curvas tracejadas indicam evoluções intermediárias entre ρ_{ξ_I} e ρ_{ξ_F} , que determinam a curva β ligando ρ_{ξ_I} a ρ_{ξ_F}	112
Figura 14 – Representação gráfica da evolução dinâmica de ρ_0 no tempo. Em azul, temos o estado evolutivo de ρ_0 no tempo $t \in [0, \tau]$. Inicialmente, a curva se aproxima daquela que fornece a evolução ρ_{ξ_I} e, conforme t chega perto de τ , a curva azul torna-se próxima de ρ_{ξ_F}	114
Figura 15 – Representação gráfica da esfera de Bloch no caso de um canal de defasagem paralela. O canal contrai a esfera mantendo a direção z intacta, assim como o centro da esfera deformada, fato que fornece a esse canal a adjetivação unital	119
Figura 16 – Representação gráfica da esfera de Bloch no caso de um canal de amortecimento de amplitude. Ele desloca a esfera na direção z junto ao seu centro, sendo um canal não unital	121

SUMÁRIO

1	INTRODUÇÃO	19
1.1	Precedentes históricos	19
1.2	Apresentação do trabalho	23
1.3	Convenções e notações utilizadas	31
2	MODELOS ESTATÍSTICOS: UMA APRESENTAÇÃO NA PERSPECTIVA DA GEOMETRIA RIEMANNIANA	33
2.1	Modelos estatísticos	33
2.2	A métrica de Fisher e o tensor de Amari-Chentsov	38
2.2.1	<i>Motivação: a informação de Fisher e a desigualdade de Cramér–Rao em metrologia</i>	38
2.2.2	<i>A informação e a métrica de Fisher</i>	41
2.2.3	<i>O tensor de Amari-Chentsov</i>	44
2.3	As α -conexões e a relação entre distribuições gaussianas e a geometria hiperbólica	45
3	GEOMETRIA DA INFORMAÇÃO FINITA	51
3.1	Introdução a uma abordagem extrínseca: apresentação de $\mathcal{P}_+(\mathcal{X})$ através de álgebra de funções sobre um conjunto finito	51
3.2	A métrica de Fisher	55
3.3	Conexões mistura e exponencial	66
3.4	O tensor de Amari-Chentsov	75
3.5	As α -conexões	81
4	VARIEDADES ESTATÍSTICAS E VARIEDADES DUALÍSTICAS	87
4.1	Estruturas dualística e estatística em uma variedade diferenciável	87
4.2	As famílias exponenciais e mistura - um motivação para a geometria da informação plana	91
4.3	A dualidade em geometria da informação e as transformadas de Legendre	95
4.3.1	<i>Introdução a variedades dualmente planas</i>	95
4.3.2	<i>Uma interpretação intuitiva sobre a transformada de Legendre</i>	99
4.4	A geometria da dualidade em geometria da informação plana	101

5	APLICAÇÃO: GEOMETRIA DA INFORMAÇÃO E LIMITES QUÂNTICOS DE VELOCIDADE	107
5.1	A introdução da geometria da informação na busca por limites quânticos de velocidade	107
5.2	A obtenção de QSLs geométricos	109
5.3	Dinâmica unitária	115
5.4	Dinâmica não unitária - canal de defasagem paralela	118
5.5	Dinâmica não unitária - canal de amortecimento de amplitude	120
6	CONCLUSÃO	123
	REFERÊNCIAS	127
APÊNDICE A	DEMONSTRAÇÃO DA CARACTERIZAÇÃO PARA O TENSOR DE AMARI-CHENTSOV EM $\mathcal{P}_+(\mathcal{X})$	133

INTRODUÇÃO

1.1 Precedentes históricos

Àqueles que se questionam sobre o que viria a ser a geometria da informação, ela pode ser introduzida, em um primeiro contato, através de um panorama histórico de seu desenvolvimento, visto que ela é uma área recente quando comparada a tantas outras tradicionais e consagradas dentro e fora da geometria. De fato, a geometria da informação integra a ciência da informação, que consiste em um termo genérico para se referir a um agrupamento de campos razoavelmente correlacionados dentro de perspectivas científicas recentes, campos esses como, por exemplo, aprendizado de máquina, estatística e teoria da informação (NIELSEN, 2022). Sendo assim, os antecedentes da geometria da informação dialogam com os estudos estatísticos e se cruzam, em especial, com os desenvolvimentos da teoria da informação, que compreende, grosso modo, os sistemas de comunicação (ASH, 1990) e que surgiu a fim de lidar com problemas, essencialmente, advindos da engenharia elétrica.

No estudo dos possíveis limitantes para a compressão e a transmissão de dados por um canal¹, imaginava-se que um aumento na taxa de transmissão de informação levava a um aumento, também, na probabilidade de erro na saída do canal (COVER; THOMAS, 2006). Contudo, entre as décadas de 1940 e 1950, o matemático estadunidense Claude Shannon demonstrou que esse era um pressuposto falso (COVER; THOMAS, 2006), introduzindo os estudos do que chamou de entropia de informação, em paralelo à entropia termodinâmica e que, atualmente, em sua homenagem, recebe o nome de entropia de Shannon.

Para além da teoria de comunicação dentro da engenharia elétrica, os desenvolvimentos da teoria da informação reverberaram em outras áreas, como na economia, por meio da implementação de uma abordagem utilizando grandezas análogas à entropia para o estudo de

¹ Por completude, é importante ressaltar que estamos chamando de “canal” um sistema que é constituído de um conjunto de entrada, um conjunto de saída e uma matriz de probabilidade que expressa a probabilidade de se observar uma saída específica dada uma entrada (COVER; THOMAS, 2006).

séries de retorno em uma bolsa de valores; como na ciência da computação, através da busca pelos algoritmos mais “simples”, o que se traduz matematicamente através da minimização da complexidade de Kolmogorov e da complexidade computacional²; como na física, em que a entropia de sistemas termodinâmicos se conecta com a entropia definida dentro da teoria da informação; e como na estatística, uma vez que as principais grandezas da teoria da informação podem ser interpretadas como funcionais de distribuições de probabilidade (COVER; THOMAS, 2006).

Previamente às produções de Shannon sobre a entropia de informação, ao estudar modelos estatísticos sob a ótica da geometria riemanniana, o econometrista estadunidense Harold Hotelling e o estatístico indiano estadunidense Callyampudi Rao desenvolveram trabalhos, respectivamente, por volta de 1930 e 1945, que produziram o cerne³ do que hoje é conhecido como geometria da informação (NIELSEN, 2022). Equipando os modelos estatísticos com um métrica riemanniana específica, chamada de métrica de Fisher, cuja nomenclatura faz referência ao orientador de doutorado de Rao, Ronald Fisher, o indiano estudou a relação entre as distâncias geodésicas e problemas de teste de hipóteses (RAO, 2021; NIELSEN, 2022), derivando, assim, a desigualdade de Cramér-Rao (AY *et al.*, 2017), em voga na metrologia quântica (MAGNO, 2020). Em coordenadas, essa métrica coincide com a chamada matriz de informação de Fisher, que possui estreita relação com a entropia de Shannon (AMARI; NAGAOKA, 2000).

Ademais, ao estudar uma maneira de mensurar a distinguibilidade entre distribuições de probabilidade, o geofísico e estatístico inglês Harold Jeffreys introduziu, entre 1946 e 1948, a divergência de Kullback-Leibler, que consiste em uma maneira de realizar essa medição. Verificando, então, o comportamento dela sob aproximações locais infinitesimais, Jeffreys notou que a matriz de informação de Fisher surgia para aproximações de segunda ordem (AY *et al.*, 2017).

Já na segunda metade do século XX, houve avanços importantes dentro dessa área emergente, dando destaque aos trabalhos do matemático soviético Nikolai Chentsov⁴ que, compartilhando da noção de que os problemas estatísticos poderiam ser interpretados como problemas de decisão, repensou a abordagem geométrica até então utilizada para entender os modelos estatísticos (NIELSEN, 2022). A partir de uma palestra realizada pelo também soviético Andrey Kolmogorov em 1955, Chentsov desenvolveu uma conexão linear plana no espaço das distribuições de probabilidade sobre um conjunto finito, chamada, posteriormente, de conexão exponencial (AY *et al.*, 2017). A partir disso, munindo modelos estatísticos de uma estrutura de

² A complexidade de Kolmogorov de uma *string* de dados é definida como o comprimento do menor programa de computador binário necessário para processar a *string*, enquanto a complexidade computacional pode ser definida como o tempo de execução do programa (COVER; THOMAS, 2006).

³ Há discordâncias entre autores sobre quando, de fato, a geometria da informação se configurou como campo de estudos. Para (AY *et al.*, 2017), ela surge por volta de 1992, com a publicação em inglês de “*Information geometry of Boltzmann machines*” (AMARI; KURATA; NAGAOKA, 1992).

⁴ Devido ao seu nome ser originalmente escrito usando o alfabeto cirílico, ele pode aparecer citado como Chentsov, Čencov ou Cencov em textos redigidos utilizando o sistema de escrita latino.

variedade diferenciável, ele estudou, em linhas gerais, como certas grandezas parametrizadas variavam tanto por reparametrizações suaves quanto por certas funções especiais, chamadas de estatísticas suficientes (NIELSEN, 2022). Diz-se então que, quando não há tais variações, essas invariâncias, juntas, configuram o que é chamado de invariância estatística. Assim, durante sua pesquisa, ele descobriu a existência de um campo tensorial covariante de *rank* 3 simétrico em todas as suas entradas caracterizado pela invariância estatística (NIELSEN, 2022).

O interesse de Chentsov sobre invariâncias não se restringiu à descoberta desse campo tensorial, mas o conduziu a provar uma importante caracterização para a métrica de Fisher em um contexto de estatística sobre amostras finitas, resultado hoje conhecido como Teorema de Chentsov e publicado, junto a outros de seu resultado, em “*Statistical decision rules and optimal inference*” em 1972 (BAKHVALOV *et al.*, 1994; CHENTSOV, 2000). Futuramente, caberia ao matemático canadense L. Lorne Campbell estudar essa caracterização para a métrica de Fisher em espaços de medidas positivas sobre uma amostra finita, o que resultaria, em 1986, na publicação do artigo “*An extended Čencov characterization of the information metric*” (CAMPBELL, 1986; AY *et al.*, 2017).

Em complemento a isso, podemos comentar que, dado que o termo “geometria da informação” só foi cunhado em 1992, a área de estudos de Chentsov foi primeiramente chamada, em russo, de algo semelhante, na língua inglesa, à “*geometrostatistics*”⁵, sendo traduzido para o inglês como “*geometrical statistics*”, esse último, que pode ser transposto para o português como “estatística geométrica” (NIELSEN, 2022).

Outrossim, a pesquisa acerca da conexão exponencial também se desenvolveu por meio das análises de Bradley Efron, estatístico estadunidense que introduziu a nomenclatura “curvatura estatística” para se referir, grosso modo, à segunda forma fundamental de subfamílias de uma família estatística específica, denominada de família exponencial. Ele interpretou essas subfamílias como subvariedades riemannianas de uma variedade diferenciável munida da métrica de Fisher (AY *et al.*, 2017). Em contato com Efron, o estatístico britânico Philip Dawid propôs uma segunda conexão linear plana, atualmente chamada de conexão mistura, que, na teoria que futuramente se desenvolveu, mostrou possuir uma intrínseca ligação com a conexão exponencial, havendo uma certa relação de dualidade entre elas (NIELSEN, 2022).

Os estudos sobre dualidade dentro da geometria da informação desenvolveram-se, especialmente, a partir das empreitadas do matemático Shun-ichi Amari e do engenheiro Hiroshi Nagaoka, ambos japoneses (AY *et al.*, 2017). Primeiramente, Amari provou que, dada uma conexão linear ∇ livre de torção em uma variedade riemanniana, existe uma segunda conexão com essas mesmas características, chamada de dual de ∇ e representada por ∇^* , tal que a conexão de Levi-Civita da métrica escreve-se como a média dessas duas conexões (NIELSEN, 2022).

⁵ Optamos por não fazer a tradução do termo para o português devido ao fato de ele ser um neologismo até mesmo no russo. Ele aparece em (ЧЕНЦОВ, НН, 1972), quando Chentsov afirma que o assunto daquela monografia poderia ser descrito pela noção de “геометростатистика” (ЧЕНЦОВ, НН, 1972).

Motivados por isso, Amari e Nagaoka introduziram a α -geometria, que consiste na geometria que emerge de considerar duas conexões, $\nabla^{(\alpha)}$ e $\nabla^{(-\alpha)}$, duais uma a outra. Essa geometria foi desenvolvida tomando α como sendo um número real, contudo, autores mais recentes, como os de (AY *et al.*, 2017), restringem α apenas ao intervalo simétrico $[-1, 1]$. Convém mencionar que a α -geometria, durante o início de seu desenvolvimento, refletia um aspecto próprio dos trabalhos dessa área na época, que era a premissa de que os modelos estatísticos, como variedades, possuíam dimensão finita (AY *et al.*, 2017). Os precursores de uma geometria da informação em dimensão infinita foram Giovanni Pistone e Carlo Sempi, estatísticos italianos que, em 1995, publicaram o trabalho “*An infinite-dimensional structure on the space of all the probability measures equivalent to a given one*”, apresentando uma generalização para dimensão infinita de resultados já conhecidos no contexto finito (AY *et al.*, 2017; PISTONE; SEMPI, 1995).

Retrocedendo um pouco, antes dos resultados de Pistone e Sempi sobre geometria da informação em dimensão infinita, o matemático dinamarquês Steffen Lauritzen apresentou uma abordagem intrínseca para o estudo de variedades vinculadas a interpretações estatísticas. Introduzindo uma estrutura geométrica nos espaço das distribuições de probabilidade sobre um conjunto, essa estrutura pode ser induzida em modelos estatísticos entendendo-os como subvariedades desse espaço maior. Esse é um tratamento extrínseco, comum nos primeiros desenvolvimentos da área (AY *et al.*, 2017). Em contrapartida, em 1987, Lauritzen definiu as variedades estatísticas, que consistem em variedades diferenciáveis munidas de uma métrica riemanniana e de um campo tensorial tal qual o descoberto por Chentsov (LAURITZEN, 1987). Assim, a partir dessa definição, é possível induzir nessa variedade conexões ∇ , ∇^* duais. Reciprocamente, por meio da introdução das conexões duais, pode-se obter uma estrutura de variedade estatística para a variedade diferenciável (AY *et al.*, 2017). Dessa forma, neste tratamento intrínseco, é possível pensar uma dualidade não apenas entre conexões, mas entre abordagens; entre aquela via dualismo desenvolvida especialmente por Amari e Nagaoka e esta proposta por Lauritzen.

Portanto, em vista desta introdução histórica, é possível interpretar a geometria da informação, sumariamente, como a área da ciência da informação que aborda problemas relacionados à estatística a partir de noções advindas da geometria diferencial (NIELSEN, 2022). Como é possível observar através do que foi apresentado, essa concisa interpretação do que é esse campo possui diversos desdobramentos, de maneira que o potencial de síntese dessa definição é melhor compreendido dentro de um panorama, mesmo que pequeno, de evolução dessa área interdisciplinar.

Como complemento ao que foi exposto, é importante salientar que a métrica de Fisher exerceu função histórica importante, também, no desenvolvimento de trabalhos em segmentos como genética populacional. No estudo de distribuições gaméticas de genótipos de uma população diploide⁶, houve a busca pela métrica que deveria ser introduzida no espaço de distribuições,

⁶ Grosso modo, os gametas são as células sexuais de um organismo, carregando apenas uma cópia de

de maneira que o gradiente, com respeito a essa métrica, da aptidão média⁷ correspondesse ao que era esperado pela teoria⁸ (AKIN, 1979). Tal métrica, que consiste na métrica de Fisher, foi proposta pelo matemático iraniano Siavash Shahshahani em 1979 (AKIN, 1979). Devido a isso, é comum encontrar a métrica de Fisher sendo referida como métrica de Shahshahani no contexto de genética populacional.

Além disso, é notável observar que, anteriormente aos trabalhos de Rao e Hotelling, já se desenvolvia a geometria hiperbólica, proveniente, principalmente, das produções do russo Nikolai Lobachevsky e das empreitadas do húngaro János Bolyai para resolver o Problema das Paralelas (LEWIS, 1920). Sem uma conexão aparente entre esses dois campos de estudos, a geometria hiperbólica e a geometria da informação interrelacionam-se por meio da métrica de Fisher em espaços de distribuições normais. Neles, essa métrica coincide com a métrica hiperbólica (AMARI; NAGAOKA, 2000), de maneira que conhecidos resultados de geometria hiperbólica bidimensional podem ser vistos sob a óptica da geometria da informação.

Atualmente, diversos trabalhos têm sido desenvolvidos em geometria da informação, especialmente inspirados pelas produções acadêmicas de Nihat Ay, Hông Vãn Lê, Frank Nielsen e Amari, autor da obra pioneira “*Methods of Information Geometry*” (AMARI; NAGAOKA, 2000). Teóricos de informação quântica e pesquisadores de aprendizado de máquina e de algoritmos quânticos⁹, por exemplo, aprimoram e aplicam resultados da área, mantendo o caráter transdisciplinar que perpassa a evolução desse campo de pesquisa.

1.2 Apresentação do trabalho

O objetivo principal deste trabalho é apresentar a geometria da informação, entendendo sua origem e os principais resultados advindos de diferentes abordagens para ela e aplicações

cada cromossomo da espécie (INSTITUTE, 2023). Assim, uma população diploide é um conjunto de indivíduos de uma mesma espécie, habitando o mesmo espaço geográfico, interagindo entre si e com o meio e cujas células, majoritariamente, possuem cromossomos que se organizam em pares, chamados de cromossomos homólogos (KLUNG *et al.*, 2019). Já o genótipo de um indivíduo diz respeito à totalidade do seu material genético (ELSTON; SATAGOPAN; SUN, 2012).

⁷ A aptidão de um tipo de indivíduo é a taxa de crescimento de uma população formada por esse tipo, ou seja, ela é seu sucesso reprodutivo (BELL, 1997). Assim, a aptidão média de uma população é a média das aptidões dos indivíduos que a constitui ponderada pela frequência de aparecimento desses indivíduos nela.

⁸ Para ser mais preciso, pelo que era esperado pelo Teorema Fundamental da Seleção Natural de Fisher. O detalhamento deste ponto foge ao escopo do trabalho. Para mais detalhes, veja (AKIN, 1979).

⁹ No sexto capítulo de (AY *et al.*, 2017), há a apresentação de aplicações de geometria da informação em campos de estudos distintos. Como exemplos de trabalhos envolvendo geometria da informação sob a perspectiva da teoria de informação quântica, podemos citar (MAGNO, 2020) e (PIRES *et al.*, 2016). Ainda dentro dessa área, uma aplicação que vem ganhando notória atenção na comunidade acadêmica diz respeito ao uso de métodos de geometria da informação em correção de erros (FIUSA; PINTO; PIRES, 2023), assunto de alto interesse prático em computação quântica. Já no final de (NIELSEN, 2022), o autor expõe como a geometria da informação pode ser utilizada na abordagem típica de aprendizado de máquina para minimização da função perda.

no estudo de limites quânticos de velocidade. Assim, ele compreende desde o entendimento de modelos estatísticos sob a perspectiva geométrica até os desdobramentos das diferentes definições para uma família destacada de conexões lineares em espaços de distribuições de probabilidade: a das α -conexões. Desse modo, este trabalho não parte da pretensão de esgotar os conceitos que serão expostos, mas de motivar uma compreensão entre suas correlações e como eles atuam como conectores entre diferentes áreas do conhecimento, em especial entre a geometria riemanniana e a estatística.

Posto isso, uma vez que a geometria da informação encarrega-se de utilizar ferramentas advindas da geometria diferencial, especialmente aquelas decorrentes da geometria riemanniana, para resolver problemas de estatística, o conceito de distribuição de probabilidade é fundamental para ela. Desse modo, dado um conjunto que constitui, em linguagem estatística, um espaço de eventos, uma distribuição de probabilidade nele pode ser entendida como uma função positiva semidefinida e normalizada partindo dele e tomando valores reais. Por meio dela, é possível realizar, por exemplo, a estimativa de algum parâmetro utilizado na modelagem de um conjunto de dados experimentais. Esse é um possível contexto de inserção dos modelos estatísticos. No capítulo 2, eles serão apresentados como famílias S de distribuições de probabilidade p_ξ parametrizadas por $\phi : \xi \in \Xi \subset \mathbb{R}^n \rightarrow S$, $\xi \mapsto p_\xi$ e que possuem uma latente estrutura de variedade diferenciável. O tratamento que será dado a eles não é o mais geral, contudo é suficiente para apresentar desenvolvimentos importantes da área, sobretudo aqueles de seu início, no que configura os primeiros passos para uma interpretação intrínseca de variedades riemannianas munidas de estruturas que refletem aspectos estatísticos, englobando, assim, os principais exemplos de famílias de distribuições. Esse capítulo seguirá uma abordagem próxima à apresentada em (AMARI; NAGAOKA, 2000), porém, trazendo aspectos recentes desenvolvidos em (AY *et al.*, 2017). A escolha por esse método de apresentação encontra subsídios para além da escolha bibliográfica, uma vez que tratamentos modernos mais amplos utilizam-se de resultados de análise em variedades de Banach, sendo esse um tópico à parte do projeto, que pode ser contornado em uma primeira exposição e que não se coloca como essencial para o que será apresentado.

Dessa maneira, interpretando modelos estatísticos como variedades diferenciáveis, é possível introduzir uma métrica riemanniana neles, com o intuito, por exemplo, de medir distinguibilidade entre distribuições de probabilidade deles. Dentre as possibilidades, uma métrica destaca-se por sua relação com um objeto já conhecido dentro dos estudos de teste de hipóteses: a matriz de informação de Fisher. Aparecendo quando há o interesse de se estimar, através de um estimador $\hat{\xi}$, o valor de um parâmetro ξ referente a um conjunto de n medidas independentes e igualmente distribuídas, a matriz de informação de Fisher $G(\xi)$ em ξ compõe um limite inferior para a variância de $\hat{\xi}$, chamada de cota de Cramér-Rao, de maneira que

$$\text{Var}_\xi \left[\hat{\xi} \right] \geq \frac{\left(\mathbb{E}_\xi \left[\hat{\xi}(x) \right]' \right)^2}{nG(\xi)}$$

é denominada de desigualdade de Cramér-Rao (MAGNO, 2020). Quando $G(\xi)$ é positiva definida, ela corresponde à matriz de uma métrica riemanniana escrita na base referente ao sistema de coordenadas ϕ , sendo essa a métrica de Fisher. Em um modelo S , ela é expressa como

$$g(\xi)(X, Y) = E_{\xi} \left[\left(\frac{\partial \ell_{\xi}}{\partial X} \right) \left(\frac{\partial \ell_{\xi}}{\partial Y} \right) \right]$$

para todo $\xi \in \Xi$ e para todos $X, Y \in T_{\phi(\xi)}S$, onde $\ell_{\xi} := \ln p_{\xi}$ (AMARI; NAGAOKA, 2000). Essa é uma métrica que, no modelo $\{\mathcal{N}(\mu, \sigma)\}$ das distribuições gaussianas, induz uma consagrada geometria. Considerando esse modelo como aquele tomado sobre \mathbb{R} composto pelas distribuições da forma

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

parametrizadas por elementos do semiplano superior

$$\Xi = \{(\mu, \sigma) \mid -\infty < \mu < \infty, 0 < \sigma < \infty\},$$

a matriz de informação de Fisher de $\{\mathcal{N}(\mu, \sigma)\}$ é $g = 2g_{\mathbb{H}^2}$, onde $g_{\mathbb{H}^2}$ é a métrica hiperbólica do modelo do Plano de Lobachevsky escrita em coordenadas retangulares, de forma que a geometria do espaço das distribuições gaussianas, na perspectiva da geometria da informação, é basicamente hiperbólica bidimensional.

Ademais, a métrica de Fisher não é o único campo tensorial importante nos modelos estatísticos do ponto de vista da geometria da informação. Sua versão de *rank* 3 possui grande relevância nos estudos de Chentsov. Sendo chamado de tensor de Amari-Chentsov ou tensor 3-simétrico, ele consiste no campo tensorial 3-covariante simétrico em todas as suas entradas e dado por

$$T(\xi)(X, Y, Z) = E_{\xi} \left[\left(\frac{\partial \ell_{\xi}}{\partial X} \right) \left(\frac{\partial \ell_{\xi}}{\partial Y} \right) \left(\frac{\partial \ell_{\xi}}{\partial Z} \right) \right]$$

para todo $\xi \in \Xi$ e para todo $X, Y, Z \in T_{\phi(\xi)}S$ (AY *et al.*, 2017). Os trabalhos de Chentsov também propunham um rompimento com a noção em uso na época de que a conexão de Levi-Civita da métrica de Fisher era a mais adequada para se estudar modelos estatísticos. Essa ideia reverberou nos trabalhos de Amari e Nagaoka, que estabeleceram as α -conexões: uma família uniparamétrica de conexões lineares em S parametrizadas por $\alpha \in [-1, 1]$ e cujos símbolos de Christoffel no sistema de coordenadas ϕ da parametrização são dados por

$$\Gamma_{ijk}(\xi) = E_{\xi} [(\partial_i \partial_j \ell_{\xi})(\partial_k \ell_{\xi})] + \frac{1-\alpha}{2} E_{\xi} [(\partial_i \ell_{\xi})(\partial_j \ell_{\xi})(\partial_k \ell_{\xi})],$$

onde ∂_i são os campos canônicos do sistema de coordenadas (AY *et al.*, 2017). Essas conexões são livres de torção e contemplam a conexão de Levi-Civita, obtida quando $\alpha = 0$. Do ponto de vista da estatística, essas conexões permitem medir a dissimilaridade entre distribuições de probabilidade de um modelo. Assim, através de uma mudança na noção de distância riemanniana, as compreensões geométricas decorrentes de $\nabla^{(\alpha)}$ podem ser adaptadas de acordo com o modelo estatístico analisado.

Feita, então, essa apresentação sobre os modelos estatísticos, que parte de uma perspectiva intrínseca de análise geométrica deles, no capítulo 3 será exposta uma abordagem extrínseca para variedades no contexto de geometria da informação finita. Assim, os espaços de parâmetro serão considerados como mergulhados em espaços de medida de dimensão finita, de maneira que a geometria deles é definida a partir desses mergulhos. Essa abordagem, transposta ao caso em que os espaços de medida têm dimensão infinita, conduz a dificuldades matemáticas que a geometria riemanniana, por si só, não resolve facilmente, precisando aliar-se à análise funcional. Devido a isso, optamos por apresentar apenas o caso finito, uma vez que há diversos ganhos advindos de uma abordagem extrínseca já nesse caso e que devem ser destacados, como, por exemplo, o uso de estruturas majoritariamente algébricas para induzir a geometria.

Desse modo, partindo de um conjunto finito não vazio \mathcal{X} de cardinalidade $n + 1$, consideraremos

$$\begin{aligned} \mathcal{F}(\mathcal{X}) &: \text{a álgebra de funções de } \mathcal{X} \text{ a } \mathbb{R}, \\ \mathcal{S}(\mathcal{X}) &: \text{o espaço dual de } \mathcal{F}(\mathcal{X}), \\ \mathcal{B}_{\mathcal{F}(\mathcal{X})} &: \text{a base canônica de } \mathcal{F}(\mathcal{X}), \text{ denotada por } \{e_1, \dots, e_{n+1}\}, \\ \mathcal{B}_{\mathcal{S}(\mathcal{X})} &: \text{a base dual de } \mathcal{B}_{\mathcal{F}(\mathcal{X})}, \text{ denotada por } \{\delta^1, \dots, \delta^{n+1}\}. \end{aligned}$$

Como os elementos de $\mathcal{S}(\mathcal{X})$ podem ser interpretados como medidas assinadas sobre a σ -álgebra discreta de \mathcal{X} , podemos definir a variedade de medidas finitas positivas sobre \mathcal{X} e a variedade de medidas de probabilidade positivas sobre \mathcal{X} , respectivamente dadas por

$$\begin{aligned} \mathcal{M}_+(\mathcal{X}) &:= \left\{ \mu = \sum_{i=1}^{n+1} \mu_i \delta^i \in \mathcal{S}(\mathcal{X}) \mid \mu_i > 0, i = 1, \dots, n+1 \right\}, \\ \mathcal{P}_+(\mathcal{X}) &:= \left\{ \xi = \sum_{i=1}^{n+1} \xi_i \delta^i \in \mathcal{M}_+(\mathcal{X}) \mid \sum_{i=1}^{n+1} \xi_i = 1 \right\}. \end{aligned}$$

Assim, em $\mathcal{M}_+(\mathcal{X})$, é plausível definirmos a métrica de Fisher em $\mu = \sum_{i=1}^{n+1} \mu_i \delta^i \in \mathcal{M}_+(\mathcal{X})$ como

$$g(\mu)(A, B) := \sum_{i=1}^{n+1} \frac{a_i}{\mu_i} \frac{b_i}{\mu_i} \mu_i$$

para todo $A = (\mu, \sum_{i=1}^{n+1} a_i \delta^i)$, $B = (\mu, \sum_{i=1}^{n+1} b_i \delta^i) \in T_\mu \mathcal{M}_+(\mathcal{X}) = \{\mu\} \times \mathcal{S}(\mathcal{X})$, onde a última igualdade é a menos de isomorfismos (AY *et al.*, 2017). Já a métrica de Fisher em $\mathcal{P}_+(\mathcal{X})$ é definida como a métrica induzida de $(\mathcal{M}_+(\mathcal{X}), g)$ na subvariedade $\mathcal{P}_+(\mathcal{X})$, métrica essa também denotada por g .

Um ponto a se destacar é que $(\mathcal{P}_+(\mathcal{X}), g)$ é isométrico ao setor esférico

$$\mathcal{S}_{2,+}(\mathcal{X}) := \left\{ f \in \mathcal{F}(\mathcal{X}) \mid f(x_i) > 0, i = 1, 2, \dots, n+1, \sum_{i \in I} f(x_i)^2 = 4 \right\} \subset \mathcal{F}(\mathcal{X})$$

munido da métrica redonda, ou seja, da métrica induzida pela métrica euclidiana em $\mathcal{F}(\mathcal{X})$ (AY *et al.*, 2017). Além disso, mostraremos que essa métrica é a única (a menos de redimensionamento) preservada por morfismos de Markov induzidos por núcleos de Markov congruentes, ou seja e grosso modo, por mapas da forma

$$K_* : \mathcal{P}(\mathcal{X}) \rightarrow \mathcal{P}(\mathcal{X}'),$$

com \mathcal{X} e \mathcal{X}' sendo conjuntos finitos não vazios tais que $|\mathcal{X}| \leq |\mathcal{X}'|$ e

$$\mathcal{P}(\mathcal{X}) := \left\{ \xi \in \mathcal{S}(\mathcal{X}) \mid \xi_i \geq 0, \sum_{i=1}^{n+1} \xi_i = 1 \right\}$$

sendo o conjunto das medidas não negativas e normalizadas sobre \mathcal{X} (AY *et al.*, 2017). Esse resultado é conhecido como Teorema de Caracterização de Chentsov, por vezes descrito como a métrica de Fisher ser a única métrica em $\mathcal{P}_+(\mathcal{X})$ preservada por estatísticas chamadas de suficientes.

Ademais, duas conexões lineares destacam-se, em especial, no contexto de geometria da informação finita. São elas a conexão mistura $\tilde{\nabla}^{(m)}$ e a conexão exponencial $\tilde{\nabla}^{(e)}$. Dados dois campos de vetores A, B em $\mathcal{M}_+(\mathcal{X})$ e dada $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathcal{M}_+(\mathcal{X})$ uma curva suave tal que $\gamma(0) = \mu$ e $\dot{\gamma}(0) = a_\mu$, com $A(\mu) = (\mu, a_\mu)$, essas conexões são definidas em termos de

$$\frac{\partial b}{\partial a_\mu}(\mu) := \lim_{t \rightarrow 0} \frac{1}{t} (b_{\gamma(t)} - b_\mu).$$

Assim, sendo $\mathfrak{X}(\mathcal{M}_+(\mathcal{X}))$ o conjunto dos campos de vetores suaves em $\mathcal{M}_+(\mathcal{X})$,

$$\begin{aligned} \tilde{\nabla}^{(m,e)} : \mathfrak{X}(\mathcal{M}_+(\mathcal{X})) \times \mathfrak{X}(\mathcal{M}_+(\mathcal{X})) &\longrightarrow \mathfrak{X}(\mathcal{M}_+(\mathcal{X})) \\ (A, B) &\longmapsto \tilde{\nabla}_A^{(m,e)} B, \end{aligned}$$

onde

$$\tilde{\nabla}_A^{(m)} B(\mu) = \left(\mu, \frac{\partial b}{\partial a_\mu}(\mu) \right), \quad \tilde{\nabla}_A^{(e)} B(\mu) = \left(\mu, \frac{\partial b}{\partial a_\mu}(\mu) - \sum_{i=1}^{n+1} \frac{a_{\mu i} b_{\mu i}}{\mu_i} \right).$$

Essas conexões, por mais que estejam definidas em $\mathcal{M}_+(\mathcal{X})$, podem ser induzidas em $\mathcal{P}_+(\mathcal{X})$, produzindo as conexões lineares $\nabla^{(m)}$ e $\nabla^{(e)}$ (AY *et al.*, 2017). Além disso, elas podem ser utilizadas para definir o tensor de Amari-Chentsov tanto em $\mathcal{M}_+(\mathcal{X})$ quanto em $\mathcal{P}_+(\mathcal{X})$, sendo esse uma versão de *rank* 3 da métrica de Fisher e definido, neste contexto, como

$$T(\eta)(A, B, C) := g(\eta) \left(\widehat{\nabla}_A^{(m)} B(\eta) - \widehat{\nabla}_A^{(e)} B(\eta), C(\eta) \right),$$

no caso de $\mathcal{M}_+(\mathcal{X})$, para todo $\eta \in \mathcal{M}_+(\mathcal{X})$, para todo $A, B, C \in \mathfrak{X}(\mathcal{M}_+(\mathcal{X}))$ e $\widehat{\nabla} = \tilde{\nabla}$ e, no caso de $\mathcal{P}_+(\mathcal{X})$, para todo $\eta \in \mathcal{P}_+(\mathcal{X})$, para todo $A, B, C \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$ e $\widehat{\nabla} = \nabla$. Dada sua origem próxima da métrica de Fisher, o tensor de Amari-Chentsov apresenta a mesma invariância por morfismo de Markov que ela, uma característica que não é observada para campos tensoriais n -covariantes em $\mathcal{P}_+(\mathcal{X})$ para $n \geq 4$.

As conexões mistura e exponencial, para além de permitirem definir o tensor de Amari-Chentsov, ao mesmo tempo que definem uma família uniparamétrica de conexões lineares em $\mathcal{M}_+(\mathcal{X})$ e em $\mathcal{P}_+(\mathcal{X})$, elas se colocam como casos especiais dela. Chamada de família das α -conexões, os elementos que a compõem são da forma

$$\widehat{\nabla}^{(\alpha)} := \frac{1-\alpha}{2} \widehat{\nabla}^{(m)} + \frac{1+\alpha}{2} \widehat{\nabla}^{(e)},$$

onde $\alpha \in [-1, 1]$ e

$$\widehat{\nabla}^{(\alpha)} = \begin{cases} \widetilde{\nabla}^{(\alpha)} & \text{em } \mathcal{M}_+(\mathcal{X}), \\ \nabla^{(\alpha)} & \text{em } \mathcal{P}_+(\mathcal{X}). \end{cases}$$

Desse modo, é possível observar que $\widehat{\nabla}^{(-1,1)} = \widehat{\nabla}^{(m,e)}$. Assim, da mesma maneira que concluímos o capítulo 2 tratando de uma abordagem intrínseca para os modelos estatísticos, finalizaremos o capítulo 3 estudando a geometria que emerge dessas α -conexões em geometria da informação finita dentro de um tratamento extrínseco.

Já no capítulo 4, seguindo as ideias desenvolvidas, especialmente, por Amari, Nagaoka e Lauritzen, estudaremos a geometria da informação partindo, não necessariamente, de um modelo estatístico, mas já de um objeto geométrico: uma variedade riemanniana. Podendo ser entendido como um complemento a 2, a escolha de apresentar esse capítulo como o último deste trabalho sustenta-se em uma construção do texto baseada na profundidade matemática dos resultados presentes em cada capítulo, sempre havendo o cuidado de manter uma coerência entre eles. Dessa maneira, em 4, dedicaremos nossa atenção às conexões duais em relação a uma métrica riemanniana g em uma variedade diferenciável M . Grosso modo, sua definição pode ser entendida como um enfraquecimento da noção de compatibilidade com a métrica g , de maneira que ∇ e ∇^* são ditas duais com respeito a g se, para todo $A, B, C \in \mathfrak{X}(M)$,

$$Cg(A, B) = g(\nabla_C A, B) + g(A, \nabla_C^* B).$$

Então, (g, ∇, ∇^*) é chamada de uma estrutura dualística em M (NIELSEN, 2020). Para além disso, as conexões ∇ e ∇^* , quando livres de torção, permitem a definição de um campo tensorial 3-covariante simétrico em todas as suas entradas e dado por

$$T := g(\nabla^* - \nabla, \cdot),$$

de maneira que (g, T) é chamado de estrutura estatística em M (AY *et al.*, 2017). Da forma como foi colocado, entende-se que (g, ∇, ∇^*) , com ∇, ∇^* livres de torção, induz uma estrutura estatística, contudo, mostraremos que a recíproca também é verdadeira: (g, T) produz uma estrutura dualística cujas conexões duais são livres de torção. Mais do que isso, ela define uma família de conexões lineares na variedade parametrizadas por $\alpha \in [-1, 1]$, que são as já citadas α -conexões, entretanto, definidas para g sendo uma métrica riemanniana arbitrária. Essas α -conexões são escritas, em termos de (g, T) e para todo $A, B, C \in \mathfrak{X}(M)$, como

$$\nabla_A^{(\alpha)} B := \nabla_A^{(0)} B - \frac{\alpha}{2} \mathcal{H}(A, B), \quad g(\mathcal{H}(A, B), C) = T(A, B, C),$$

onde $\nabla^{(0)}$ é a conexão de Levi-Civita de g .

Apresentadas, então, as α -conexões definidas para uma métrica riemanniana qualquer, passaremos, por fim, a estudar, em 4, a geometria da informação plana. Para motivar tal estudo, veremos duas classes de famílias de distribuições de probabilidade: a das exponenciais e a das mistura. Sendo, respectivamente, $\nabla^{(1)}$ e $\nabla^{(-1)}$ conexões planas nelas, a geometria dessas famílias, vista quando g é a métrica de Fisher, pode ser descrita em termos de estruturas hessianas, isto é, de uma conexão plana e de uma métrica que, localmente, escreve-se como a hessiana de uma função suave. Assim, por meio dessa motivação, mostraremos que, dada (g, ∇, ∇^*) uma estrutura dualmente plana em M , ou seja, uma métrica riemanniana em M junto a duas conexões planas duais, e, dado um sistema de coordenadas afim para cada uma das conexões, é possível encontrar duas funções estritamente convexas $\psi, \varphi : M \rightarrow \mathbb{R}$ tais que g e g^{-1} escrevem-se como a hessiana dessas funções em relação aos sistemas afim. Na realidade, ψ e φ conectam-se através de uma transformada de Legendre que, intuitivamente, descreve o formato de cada uma dessas função em termos da outra e dos sistemas de coordenadas. Assim, uma estrutura dualmente plana (g, ∇, ∇^*) em M define uma função ψ com as propriedades que foram citadas. Por outro lado, mostraremos que a contrapartida também é verdadeira, de forma que uma (g, ∇, ∇^*) é, localmente, equivalente à informação de uma única função convexa ψ , de maneira que a geometria que emerge da estrutura dualmente plana (g, ∇, ∇^*) pode ser observada em termos de ψ .

Por fim, no capítulo 5, seguindo o que é exposto no artigo de revisão (PIRES *et al.*, 2016), apresentaremos uma aplicação da geometria da informação na busca por limites quânticos de velocidade, referidos pela sigla QSLs e que são interpretados como princípios de incerteza para o tempo e a energia dentro do contexto de sistemas quânticos. Assim, no estudo de distinguibilidade entre dois estados de um sistema através do uso de modelos estatísticos, a inserção de uma métrica riemanniana nesses modelos conduz à utilização de comprimentos geodésicos com respeito à conexão de Levi-Civita desse métrica para produzir QSLs geométricos. Dessa maneira, considerando $\mathcal{D}(\mathcal{H})$ o espaço de operadores densidade sobre o espaço de Hilbert \mathcal{H} associado a um dado sistema quântico, existe uma família de métricas riemannianas g^f em $\mathcal{D}(\mathcal{H})$ cujos elementos apresentam propriedades físicas desejáveis para distinguibilidade de densidades. Associadas univocamente a funções $f : (0, +\infty) \rightarrow (0, +\infty)$ específicas, chamadas de funções de Morozova-Chentsov, essas métricas são contrativas sob mapas semipositivos que preservam traço. Entre elas, encontra-se a métrica de Fisher quântica, que consiste em um análogo quântico da métrica de Fisher clássica apresentada no capítulo 2. Além disso, os elementos de comprimento quadrático induzidos por essas métricas entre dois operadores densidade vizinhos ρ e $\rho + d\rho$ são da forma

$$ds^2 = \frac{1}{4} \left[\sum_{j=1}^n \frac{(\langle j|d\rho|j\rangle)^2}{p_j} + 2 \sum_{\substack{j,l=1 \\ j < l}}^n \frac{1}{p_l f(p_j/p_l)} |\langle j|d\rho|l\rangle|^2 \right],$$

onde $\rho = \sum_{j=1}^n p_j |j\rangle\langle j|$, com $0 < p_j \leq 1$ para todo $j = 1, \dots, n$ e $\sum_{j=1}^n p_j = 1$, é a decomposição

espectral de ρ e onde facilmente identifica-se o primeiro termo dentro do colchete como a informação de Fisher clássica em ρ .

Então, sendo ξ um r -parâmetro que determina a evolução ρ_ξ de ρ_0 , dadas duas dinâmicas ρ_{ξ_I} e ρ_{ξ_F} dessa forma, podemos considerar uma mudança analítica do parâmetro ξ_I para ξ_F , determinando, assim, uma curva α em $\mathcal{D}(\mathcal{H})$ ligando ρ_{ξ_I} e ρ_{ξ_F} . Sendo, então, τ o tempo de evolução de ρ_0 , podemos tomar a parametrização $t \in [0, \tau] \mapsto \xi(t)$ tal que $\xi(0) = \xi_I$ e $\xi(\tau) = \xi_F$. Com $\mathcal{D}(\mathcal{H})$ munida de uma métrica g^f sob a hipótese de que $(\mathcal{D}(\mathcal{H}), g^f)$ é geodesicamente completa, o comprimento $\mathcal{L}^f(\rho_0, \rho_\tau)$ da geodésica minimizante ligando ρ_0 ao estado final ρ_τ é um limitante inferior para o comprimento $\ell_\alpha^f(\rho_0, \rho_\tau)$ da curva α de evolução do estado ρ_0 a ρ_τ , isto é,

$$\mathcal{L}^f(\rho_0, \rho_\tau) \leq \ell_\alpha^f(\rho_0, \rho_\tau).$$

Essa expressão pode ser interpretada como um QSL geométrico, contudo, o cálculo de comprimentos geodésicos é um processo difícil em diversos casos. Suas expressões são conhecidas apenas para o caso da métrica de Fisher quântica e da métrica de Wigner-Yanase, essa última obtida por meio da função de Morozova-Chentsov

$$f_{WY}(t) = \frac{(\sqrt{t} + 1)^2}{4}.$$

Assim, estudaremos a qualidade dos QSLs gerados por essas duas métricas em três dinâmicas: unitária, não unitária com um canal de defasagem paralela e não unitária com um canal de amortecimento de amplitude.

No primeiro caso, quando nos atemos às situações em que ρ_0 comuta com ρ_τ , em termos da variância média do hamiltoniano que evolui ρ_0 , a métrica de Fisher quântica g^{QF} produz QSLs mais apertados que a métrica g^{WY} de Wigner-Yanase, fenômeno que não se observa de forma tão geral no caso de sistemas abertos. Observando uma evolução tal como essa no contexto de um único *qubit*, mas na qual foi introduzido um canal de defasagem paralela, o que se nota é que, em regimes de baixa frequência, g^{WY} produz QSLs tão melhores que os obtidos por g^{QF} quanto os estados iniciais estejam próximos de $|0\rangle$ e $|1\rangle$ na esfera de Bloch. Por fim, no caso desse mesmo *qubit*, mas trocando esse canal por um canal de amortecimento de amplitude, os QSLs obtidos por meio da métrica de Wigner-Yanase são quase saturados em toda a esfera de Bloch, enquanto que os QSLs advindo da métrica de Fisher quântica quase se saturam apenas nas proximidades de $|0\rangle$. Logo, o análogo quântico da métrica de Fisher clássica mostra-se como um bom gerador de QSLs quando comparado com g^{WY} no caso unitário, enquanto que sistemas mais realísticos priorizam a métrica de Wigner-Yanase para produzir melhores limitantes para o tempo de evolução τ .

É importante pontuar que a escrita deste trabalho pressupõe conhecimentos básicos de geometria riemanniana, estatística e teoria da medida. Quando for considerado necessário apresentar alguma definição ou resultado dessas áreas, isso será feito, mesmo que algumas vezes

em notas de rodapé ou sem os demonstrar, dado que os abordar no contexto mais geral foge ao escopo deste trabalho.

1.3 Convenções e notações utilizadas

Neste trabalho, utilizamos algumas notações já bem estabelecidas nas literaturas de matemática, contudo, por completude e melhor compreensão do texto, decidimos as especificar aqui:

- $\mathbb{Z}_{\geq 0}$: conjunto dos números inteiros não negativos;
- $\mathfrak{X}(M)$: conjunto dos campos de vetores suaves na variedade diferenciável M ;
- $E_{\xi}[f]$: média de f sobre seu domínio, com ξ elemento de parametrização dela;
- $L^n(\mathcal{X})$: conjunto das funções $f: \mathcal{X} \rightarrow \mathbb{R}$ mensuráveis tais que $|f|^n$ é Lebesgue integrável sobre o conjunto¹⁰ \mathcal{X} , onde $n \in [1, +\infty)$.

Além disso, diversas vezes aparecerão os conjuntos $\{1, 2, \dots, n\}$ e $\{1, 2, \dots, n+1\}$, de maneira que é razoável os denotar por J e I , respectivamente, a fim de manter a notação mais sucinta. Ainda sobre questões de nomenclatura, durante este trabalho, quando dissermos que uma conexão linear está definida em uma variedade M , entende-se que ela está definida no fibrado tangente de M . Já tratando de tensores, não faremos uso de espaçamentos vazios nos subíndices e nos superíndices, de maneira que, por exemplo, $T_{i k}^j$ será escrito como T_{ik}^j .

Por fim, é importante pontuar que todas as funções que serão utilizadas no decorrer dos desenvolvimentos aqui presentes são suaves e mensuráveis, mesmo que isso não seja citado ou que a σ -álgebra esteja subentendida.

¹⁰ Por mais que L^n seja sucintamente apresentado assim, formalmente, dado (\mathcal{X}, X, μ) espaço de medida e $n \in [1, +\infty)$, $L^n(\mathcal{X})$ é o conjunto de todas as classes de μ -equivalência de funções X -mensuráveis a valores reais f que têm integral de $|f|^n$ finita sobre X e com respeito a μ (BARTLE, 1995).

MODELOS ESTATÍSTICOS: UMA APRESENTAÇÃO NA PERSPECTIVA DA GEOMETRIA RIEMANNIANA

2.1 Modelos estatísticos

Historicamente, o caráter probabilístico de determinados experimentos foi objeto de discussão entre cientistas. Suponha a realização de um experimento em que se obtém um conjunto finito de dados a partir do qual se quer averiguar certos aspectos de uma teoria. Devido aos limites experimentais impostos pela realidade, a análise desses dados utiliza-se de maneira central do conceito de probabilidade. Contudo, os debates desse caráter probabilístico não se restringiram apenas à forma de realização do experimento, mas à própria teoria. Exemplo disso foram as discordâncias existentes durante a unificação das descrições corpuscular e ondulatória da matéria e da radiação, nos primórdios da mecânica quântica, no que remonta à primeira metade do século XX. Enquanto físicos como Albert Einstein e Louis de Broglie, com base nas leis determinísticas da mecânica clássica, criticavam a existência de uma indeterminação fundamental na natureza, Werner Heisenberg e Niels Bohr defendiam a essencialidade da interpretação probabilística dos fenômenos quânticos ([EISBERG; RESNICK, 1979](#)).

Diante da importância de uma perspectiva probabilística dos dados, a inferência estatística os estuda partindo do pressuposto de que eles são gerados a partir de algum elemento de uma família conhecida de distribuições de probabilidade. Nesse contexto, pode-se determinar se o valor de um parâmetro da distribuição encontra-se em algum conjunto específico, tarefa que corresponde ao teste de hipóteses ([MURRAY; RICE, 2017](#)). Assim, um modelo estatístico, grosso modo entendido como essa família de distribuições, desempenha uma importante função dentro da estatística. Durante o desenvolvimento da geometria da informação e da estatística, em si, o conceito de modelo estatístico foi sendo ampliado através da introdução de novos elementos

da análise funcional. Assim, em textos mais antigos, como em (AMARI; NAGAOKA, 2000), a definição para modelo estatístico assemelha-se muito à definição de variedade diferenciável, enquanto que, em (AY *et al.*, 2017), essa definição é feita empregando noções de análise em variedades de Banach. Desse modo, por simplicidade, trataremos os modelos estatísticos de maneira análoga àquela apresentada em (AMARI; NAGAOKA, 2000), dado que essa abordagem será suficiente para o que faremos.

Pontuado isso, vamos introduzir de maneira mais formal o conceito de distribuição de probabilidade, que é central neste trabalho.

Definição 1 (Distribuição de probabilidade). (AMARI; NAGAOKA, 2000) Para $n \in \mathbb{Z}_{\geq 0}$, seja \mathcal{X} um subconjunto não vazio de \mathbb{R}^n ou um conjunto discreto não vazio de cardinalidade finita ou infinita contável. Definimos uma distribuição de probabilidade em \mathcal{X} como uma função $p: \mathcal{X} \rightarrow \mathbb{R}$ que é positiva semidefinida e normalizada em \mathcal{X} , ou seja,

1. $p(x) \geq 0$ para todo $x \in \mathcal{X}$,
2. $\int_{\mathcal{X}} p(x) dx = 1$,

onde a integral acima se reduz ao somatório $\sum_{x \in \mathcal{X}} p(x) = 1$ no caso discreto.

Note que, nesta definição, está implícita a solicitação de que p seja Lebesgue integrável. Sob a perspectiva de teoria da medida, uma distribuição p corresponde à derivada de Radon-Nikodým $\frac{dP}{d\mu}$ de uma medida de probabilidade P no espaço mensurável $(\mathcal{X}, \mathcal{B})$ com respeito à medida μ de Lebesgue ou de contagem nesse mesmo espaço¹, onde \mathcal{B} é a σ -álgebra de Borel² em \mathcal{X} . Além disso, no caso de \mathcal{X} discreto, p é usualmente chamada de um função probabilidade em \mathcal{X} , enquanto que no caso de \mathcal{X} contínuo, ela costuma ser referida como uma função densidade de probabilidade em \mathcal{X} . A fim de fazer uma abordagem mais ampla, usaremos o termo “distribuição de probabilidade” com o intuito de englobar ambos os casos. Com isso, vamos definir o que é um modelo estatístico em \mathcal{X} .

Definição 2 (Modelo estatístico). (AMARI; NAGAOKA, 2000) Seja S uma família de distribuições de probabilidade em \mathcal{X} e suponha que cada uma delas possa ser parametrizada por n variáveis a valores reais, digamos (ξ_1, \dots, ξ_n) , de forma que³

$$S = \{p_{\xi} = p(x; \xi) \mid \xi = (\xi_1, \dots, \xi_n) \in \Xi\},$$

¹ Usualmente, considera-se a medida de Lebesgue quando \mathcal{X} é um subconjunto não vazio de \mathbb{R}^n e a medida de contagem quando \mathcal{X} é um conjunto discreto não vazio. Fixamos essas duas pois são as mais usuais nesse contexto, porém μ poderia ser uma medida σ -finita qualquer em $(\mathcal{X}, \mathcal{B})$ (AMARI; NAGAOKA, 2000).

² Se \mathcal{X} for discreto, é a σ -álgebra gerada pela topologia discreta em \mathcal{X} . Já, se $\mathcal{X} = \mathbb{R}^n$, $n \in \mathbb{Z}_{\geq 0}$, ela corresponde à σ -álgebra gerada pela topologia canônica de \mathbb{R}^n .

³ $\xi = (\xi_1, \dots, \xi_n)$ pode ser chamado um parâmetro n -dimensional para S .

onde Ξ é subconjunto aberto de \mathbb{R}^n e o mapa $\phi : \Xi \rightarrow S$ dado por $\phi(\xi) = p_\xi$ é injetivo e suave em todo o seu domínio⁴. Chamamos Ξ um espaço de parâmetros e S um modelo estatístico n -dimensional.

Assim, quando estiverem subentendidos tanto o espaço de parâmetros quanto o conjunto \mathcal{X} , os elementos de S serão representados apenas por p_ξ . Além disso, usualmente, não se solicita nem a suavidade de ϕ nem que Ξ esteja na topologia de \mathbb{R}^n , contudo, aqui isso foi exigido para que seja possível utilizar resultados de geometria diferencial de maneira mais geral, sem que haja preocupação quanto a dificuldades para a diferenciação de funções.

Outrossim, o que está por trás da definição 2 é o fato de que estamos interpretando os elementos do espaço de parâmetros Ξ como medidas de probabilidade estritamente positivas em \mathcal{X} . É importante pontuar que alguns exemplos típicos de famílias de distribuições inserem-se como exemplos de modelos estatísticos, modelos esses relevantes na inferência estatística.

Para ilustrar, podemos estudar primeiramente o caso de \mathcal{X} finito. Seja, então, \mathcal{X} um conjunto finito não vazio. Escrevendo $\mathcal{X} = \{x_1, x_2, \dots, x_{n+1}\}$, usualmente um modelo para \mathcal{X} é apresentado como o conjunto dos $p(x_i; \xi) : \mathcal{X} \times \Xi \rightarrow \mathbb{R}$, $i \in I$, dados por

$$p(x_i; \xi) = \begin{cases} \xi^i, & \text{se } 1 \leq i \leq n, \\ 1 - \sum_{i \in J} \xi^i, & \text{se } i = n+1, \end{cases}$$

onde o aberto $\Xi \in \mathbb{R}^n$ é

$$\Xi = \left\{ (\xi^1, \dots, \xi^n) \in \mathbb{R}^n \mid \xi^i > 0, i \in J, \sum_{i \in J} \xi^i < 1 \right\}, \quad (2.1)$$

ou seja, Ξ é um n -simplexo, por vezes chamado de simplexo de probabilidade⁵, e esse é um modelo n -dimensional (AMARI; NAGAOKA, 2000).

Outro exemplo que podemos citar é a família de distribuição de Poisson, cujos elementos são distribuições sobre conjuntos discretos. Em outras palavras, nesse caso, $\mathcal{X} = \mathbb{Z}_{\geq 0}$, $\Xi = \{\xi \in \mathbb{R} \mid \xi > 0\}$ e os elementos desse modelo são da forma

$$p(x; \xi) = e^{-\xi} \frac{\xi^x}{x!},$$

⁴ Por suave entende-se que, para todo $x \in \mathcal{X}$, a função $\xi \mapsto p(x; \xi)$ é suave como função de \mathcal{X} em \mathbb{R} (AMARI; NAGAOKA, 2000).

⁵ Efetivamente, um n -simplexo de probabilidade é um conjunto da forma

$$\left\{ (x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_i \geq 0, i \in I, \sum_{i \in I} x_i = 1 \right\},$$

também chamado de um n -simplexo unitário padrão (NIELSEN, 2018; DOUVEN, 2022). Contudo, como a soma $\sum_{i \in I} x_i$ em 2.1 tem valor máximo 1, a troca da igualdade pela desigualdade nessa soma produz um $(n+1)$ -simplexo.

de maneira que a dimensão dele é unitária. O parâmetro ξ recebe o nome de “taxa de ocorrência” ou, simplesmente, média ou esperança, uma vez que o valor esperado de uma variável aleatória sujeita à distribuição de Poisson coincide com ξ . As distribuições desse tipo são utilizadas usualmente para estudar eventos que acontecem de maneira aleatória no espaço ou no tempo, ou seja, cuja determinação local ou temporal da ocorrência dos eventos é aleatória e eles são independentes entre si (BLOM, 1989). Assim, esse modelo é costumeiramente aplicado para o estudo de decaimentos radioativos e erros de digitação por página de um livro.

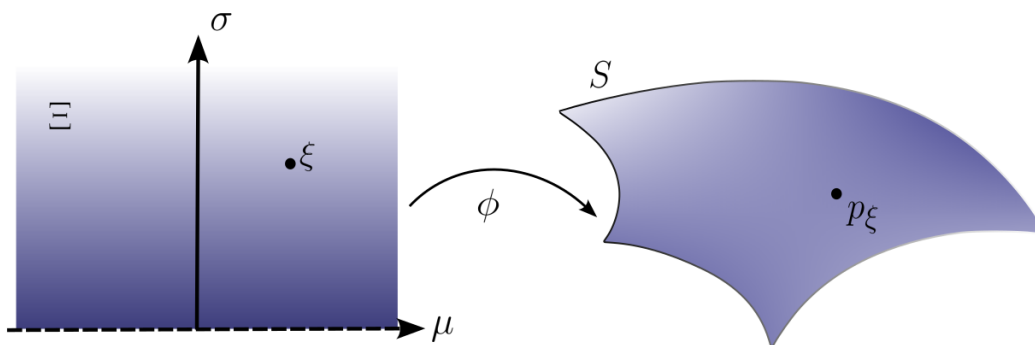
Mais um exemplo de modelo estatístico é aquele cujas distribuições são gaussianas, também chamadas de normais. Nesse caso, $\mathcal{X} = \mathbb{R}$, o espaço de parâmetros $\xi = (\mu, \sigma)$ é dado por

$$\Xi = \{(\mu, \sigma) \mid -\infty < \mu < \infty, 0 < \sigma < \infty\}$$

e as distribuições são da forma

$$p(x; \xi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.$$

Figura 1 – Representação gráfica de um modelo estatístico gaussiano. Cada elemento do semiplano euclidiano superior pode ser observado como uma distribuição gaussiana



Fonte: Elaborada pelo autor.

Os parâmetros μ e σ são, respectivamente, a média e o desvio padrão da distribuição, que usualmente é denotada por $\mathcal{N}(\mu, \sigma)$. Além disso, as distribuições gaussianas são aproximações contínuas para distribuições de Poisson com $\mu = \sigma = \xi_p$, onde $\xi_p \gg 1$ representa o ξ da poissoniana que se quer aproximar (BLOM, 1989). É importante ressaltar, também, que a relevância dessa família de distribuições pode ser vista a partir do Teorema do Limite Central que pode ser enunciado como está abaixo.

Teorema 1 (Teorema do Limite Central de Liapunov). (MAGALHÃES, 2011) Seja $X = \{X_n \mid n \geq 1\}$ uma sequência de variáveis aleatórias independentes, com média μ_n e variância σ_n^2 , com $\sigma_n^2 < \infty$ e pelo menos um dos σ_n^2 's maior do que zero. Sejam $S_n = X_1 + X_2 + \dots + X_n$ e $s_n^2 = \sum_{k \in J} \sigma_k^2$. Se a condição de Liapunov estiver satisfeita, isto é, se existir $\delta > 0$ tal que

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{k \in J} E \left[|X_k - \mu_k|^{2+\delta} \right] = 0,$$

então

$$Z_n := \frac{S_n - \mathbb{E}[S_n]}{s_n}$$

tem distribuição que tende à distribuição gaussiana com média 0 e variância 1.

Esse teorema expressa uma característica importantes das distribuições normais: desde que a condição de Liapunov seja satisfeita, não importando qual seja a distribuição das variáveis independentes que estão sendo analisadas, para amostras grandes o suficiente, a distribuição de Z_n se aproxima satisfatoriamente de uma distribuição gaussiana. Grosso modo, esse teorema permite reduzir o estudo das distribuições de probabilidade dos elementos de X , para amostras suficientemente grandes, a uma distribuição de dois parâmetros, desde que as variáveis sejam independentes entre si e que a convergência da distribuição de Z_n ocorra. Sob a óptica da estatística, ele ratifica a importância da definição 2 como ela foi apresentada.

A partir de aqui, façamos duas suposições. A primeira consiste em admitir que as derivadas $\partial_i := \frac{\partial}{\partial \xi_i}$ podem ser trocadas com as integrais de p em \mathcal{X} , de forma que vale⁶

$$\int_{\mathcal{X}} \partial_i p(x; \xi) dx = \partial_i \int_{\mathcal{X}} p(x; \xi) dx = \partial_i 1 = 0. \quad (2.2)$$

Essa suposição pode ser entendida na perspectiva da estimativa de parâmetros, significando, sem grande rigor, que a função escolhida para estimar um parâmetro de um dado modelo, na média, aproxima-se suficientemente do valor verdadeiro dele. Essa colocação ficará mais clara na seção 2.2.1, quando a métrica de Fisher for motivada a partir da metrologia.

Já, aspirando garantir a existência da métrica de Fisher em S , vamos pressupor que, para todo $p_\xi \in S$ e para todo $\xi \in \Xi$, o suporte $\text{supp}(p_\xi) := \{(x, \xi) \mid p(x; \xi) > 0\}$ de p_ξ é constante com respeito a ξ . Redefinindo \mathcal{X} como $\text{supp}(p) := \{x \in \mathcal{X} \mid p(x) > 0\}$, obtemos que S é simplesmente um subconjunto de $\mathcal{P}_+(\mathcal{X})$ definido como

$$\mathcal{P}_+(\mathcal{X}) := \left\{ p : \mathcal{X} \rightarrow \mathbb{R} \mid p(x) > 0 \text{ para todo } x \in \mathcal{X}, \int_{\mathcal{X}} p(x) dx = 1 \right\}.$$

Observando a definição de modelo estatístico, é possível perceber que S apresenta estrutura natural de variedade diferenciável, no sentido de que a estrutura diferenciável \mathcal{A} dela é constituída unicamente pela carta global (S, ϕ^{-1}) , ou seja, $\mathcal{A} = \{(S, \phi^{-1})\}$. Dessa forma, ϕ^{-1} é um sistema de coordenadas em S , cujos campos canônicos $\frac{\partial}{\partial \xi_i}$ serão denotados por ∂_i para $i \in J$. Assim, quando estivermos tratando de modelos estatísticos, estaremos os considerando como variedades diferenciáveis. Observe que podemos mudar os parâmetros de S através de um difeomorfismo $\psi : \tilde{\Xi} \rightarrow \Xi$, com $\tilde{\Xi} \subset \mathbb{R}^n$ também aberto. Isso produz um segundo modelo

⁶ Quando uma estrutura de variedade for adicionada aos modelos estatísticos, isso significará que, dado um sistema de coordenadas no modelo, seus campos coordenados ∂_i , agindo como derivações em p_ξ , podem ser trocados com as integrais de p em \mathcal{X} .

estatístico \tilde{S} com atlas natural $\tilde{\mathcal{A}} = \{(\tilde{S}, (\phi \circ \psi)^{-1})\}$. Note que, essencialmente, S e \tilde{S} codificam a mesma geometria, apesar de serem modelos distintos. Isso motivou autores como Amari a considerar, em (AMARI; NAGAOKA, 2000), como equivalentes modelos como S e \tilde{S} , que se diferem apenas por uma mudança de parâmetros $\psi : \tilde{\Xi} \rightarrow \Xi$ que é um difeomorfismo⁷.

2.2 A métrica de Fisher e o tensor de Amari-Chentsov

2.2.1 *Motivação: a informação de Fisher e a desigualdade de Cramér–Rao em metrologia*

A fim de motivar a definição de métrica de Fisher a partir da informação de Fisher, partiremos de uma construção da desigualdade de Cramér–Rao tomando como contexto a metrologia. Esta subseção é baseada em (MAGNO, 2020; ZHENG, 2011), de maneira que acompanharemos de perto alguns detalhes presentes nessas referências.

Suponha que tenhamos um conjunto de dados de um experimento, a partir do qual queremos estimar algum parâmetro específico. Esse é um empreendimento que se coloca dentro dos estudos da metrologia, que se configura como a ciência que compreende os aspectos teóricos e práticos da medição (KOBAYOSHI, 2018). Assim, a partir da análise desses dados, podemos nos questionar a respeito da quantidade de informação⁸ que eles carregam sobre o parâmetro cuja estimativa é de nosso interesse. É esse questionamento que guiará o que será apresentado nesta subseção.

Isso posto, considere que desejamos estimar o valor verdadeiro ξ_v , constante, supondo sua existência, de um único parâmetro ξ referente a um fenômeno que estamos experienciando. Para tal, preparamos nosso sistema, que se encontra em um estado inicial. Depois disso, ocorre um processo dinâmico que muda esse estado inicial para um estado final. Após isso, fazemos a medição experimental e armazenamos na variável aleatória X_1 o resultado da medida. Repetimos esse procedimento $n \in \mathbb{N}$ vezes, guardando em X_i o resultado da i -ésima medida. Ao final dessas repetições, temos o vetor aleatório⁹ $X = (X_1, \dots, X_n)$.

Com o intuito de estimar ξ_v , utilizamos um estimador não enviesado de ξ . Por estimador não enviesado entende-se um estimador tal que a média sobre todas as amostras de uma estimativa é o valor verdadeiro do parâmetro, ou seja, aquele que, em termos amplos, seria obtido em um processo de medição sem erros (MURRAY; RICE, 2017). Em outras palavras, $\hat{\xi}_n(X)$ toma a

⁷ Em (AMARI; NAGAOKA, 2000), Amari nomeia de variedade estatística um modelo S munido da estrutura diferenciável dada pelo quociente sob a relação de equivalência que identifica S e \tilde{S} como equivalentes. Contudo, nos trabalhos de Lauritzen, como (LAURITZEN, 1987), essa nomenclatura aparece em um contexto distinto, sendo usada no estudo da geometria que emerge ao partirmos de uma variedade diferenciável e darmos a ela uma estrutura que reflete aspectos estatísticos.

⁸ Aqui, informação possui um sentido intuitivo e pouco preciso.

⁹ Um vetor aleatório é uma n -upla de variáveis aleatórias.

verdadeira¹⁰ distribuição de probabilidade $p_n(X; \xi_v)$ de X e estima o parâmetro ξ , de maneira que, para obtermos ξ_v , idealmente precisamos conhecer $p_n(X; \xi_v)$. Contudo, em geral, não é conhecida a verdadeira distribuição $p_n(X; \xi_v)$, nem mesmo, previamente, qual é o valor de ξ_v . Devido a isso, usualmente delimitamos os valores que $\hat{\xi}_n(X)$ pode assumir a um conjunto $\Xi \subset \mathbb{R}$, que corresponde, nesse exemplo, ao já apresentado espaço de parâmetros. Desse modo, para cada $\xi \in \Xi$, temos uma distribuição $p_n(X; \xi)$, onde, idealisticamente, $p_n(X; \xi_v)$ é alcançada quando $n \rightarrow \infty$, supondo que essa convergência ocorre.

Chamando $p_n(X; \xi)$ de função de verossimilhança, como queremos encontrar o parâmetro de Ξ mais provável, definimos o estimador que nos fornece esse parâmetro, chamado de estimador de máxima verossimilhança, dado por

$$\hat{\xi}_n^{ML}(X) := \arg \max_{\xi \in \Xi} p_n(X; \xi).$$

Lembrando que a função logaritmo é monótona crescente, tomar o máximo dela, em essência, é o mesmo que tomar o máximo de seu argumento. Desse modo, podemos reescrever $\hat{\xi}_n^{ML}(X)$ como

$$\hat{\xi}_n^{ML}(X) = \arg \max_{\xi \in \Xi} \ell_n(X; \xi), \quad \ell_n(X; \xi) := \ln p_n(X; \xi).$$

A maximização de ℓ_n em Ξ dada pela expressão anterior pode ser entendida via o cálculo da derivada de ℓ_n , que é dada por

$$\ell'_n(X; \xi) = \frac{p'_n(X; \xi)}{p_n(X; \xi)}.$$

Assim, quando $\xi \rightarrow \xi_v$,

$$\ell'_n(X; \xi) = \frac{1}{p_n(X; \xi)} \frac{\partial p_n(X; \xi)}{\partial \xi} \rightarrow 0.$$

Perceba, agora, que a ocorrência de eventos de grande probabilidade não nos proporciona, no geral, grande informação sobre o que estamos estudando pois, realizando o experimento apenas algumas vezes, conseguimos analisar consideravelmente o fenômeno de forma empírica. Em contrapartida, se um evento pouco provável ocorre, ele nos fornece muita informação, permitindo que novas hipóteses sejam feitas, por exemplo, ou que a teorização que estava sendo formulada seja questionada. Esse último caso pode ser traduzido, matematicamente, em $|\ell'_n(X; \xi)|$ ser grande quando comparado com os eventos mais prováveis ou, equivalentemente, $\ell'_n(X; \xi)^2$ ser grande sob essa mesma comparação. Em outras palavras, $\ell'_n(X; \xi)^2$ é um quantificador de informação que conseguimos obter sobre o fenômeno a partir de X .

Suponha, então, que uma única observação seja feita, ou seja, que $n = 1$. Nessa situação, X é apenas uma variável aleatória. Note que o que está sendo feito é tentar codificar a informação a respeito do experimento em termos do parâmetro ξ : quanto mais próximo ξ estiver de ξ_v , menos informações é obtida. Contudo, a expressão para $\ell'_1(X; \xi)^2$ é dependente dos eventos que

¹⁰ Com o adjetivo “verdadeiro”, queremos nos referir às grandezas referentes a ξ_v e que, idealmente, correspondem ao fenômeno estudado.

podem ocorrer, ou seja, dos valores que X pode assumir. Para lidar com isso, tomemos a média de $\ell'_1(X; \xi)^2$ sobre X . Assim, definimos

$$G(\xi) := E_\xi [\ell'_1(x; \xi)^2] = \int_{\mathcal{X}} \ell'_1(x; \xi)^2 p_1(x; \xi) dx,$$

chamada de informação de Fisher em ξ , como essa medida informacional¹¹.

Suponha, agora, que $n \geq 1$ observações sejam feitas e que X_1, X_2, \dots, X_n são independentes e igualmente distribuídas. Nesse caso, a independência nos garante que uma medição não irá influenciar nas outras, de maneira que podemos as tratar como observações individuais. Assim, a informação codificada na i -ésima variável aleatória é apenas

$$G_i(\xi) := \int_{\mathcal{X}} \ell'_i(x; \xi)^2 p_i(x; \xi) dx. \quad (2.3)$$

Já, como as variáveis aleatórias são igualmente distribuídas, sabemos que as observações são obtidas a partir da mesma distribuição de probabilidade. Dessa forma, ao tomarmos a média sobre todos os valores acessíveis a cada variável ponderados por $p_i(x; \xi)$, teremos que $p_i(x; \xi)$ será a mesma para todas essas variáveis. Logo, obteremos o mesmo valor $G_i(\xi)$ para todas elas. Desse modo, a informação fornecida por $X = (X_1, \dots, X_n)$, pensada como somatório das informações 2.3 de cada X_i , é simplesmente

$$G(\xi) := \sum_{i \in J} G_i(\xi) = nG_n(\xi).$$

Perceba que esse resultado simples advém do fato das variáveis serem independentes e igualmente distribuídas. Nos casos mais gerais, esperamos que uma medição possa interferir na outra, de forma que 2.3 dependa das outras observações, de maneira que a observação i influencia na observação j e vice-versa. Isso nos fornece indícios de que uma maneira mais geral de tratar essa informação é através de uma matriz, chamada de matriz de informação de Fisher.

Ainda tratando do caso de n medidas independentes e igualmente distribuídas, suponha que $\text{Var}_\xi [\widehat{\xi}]$ é finita. Defina $\bar{\xi}(\xi) := E_\xi [\widehat{\xi}(x)]$ e perceba que

$$E_\xi [\ell'_n(x; \xi)] = \int_{\mathcal{X}} \frac{p'_n(x; \xi)}{p_n(x; \xi)} p_n(x; \xi) dx = 0,$$

$$\text{Var}_\xi [\ell'_n(x; \xi)] = E_\xi [(\ell'_n(x; \xi))^2] - (E_\xi [\ell'_n(x; \xi)])^2 = nG_n(\xi),$$

onde a primeira expressão acima utiliza-se da suposição 2.2. Como $\ell'_n(x; \xi)$ é usado para auxiliar na determinação do valor dos parâmetros que maximizam a função de verossimilhança, é legítimo mensurar a dependência entre ele e $\widehat{\xi}(x)$, o que pode ser feito através do cálculo da covariância

¹¹ Quando estivermos tratando dos possíveis valores assumidos por X , usaremos x .

entre eles. Desse modo, temos que

$$\begin{aligned} \text{Cov}_\xi \left[\widehat{\xi}(x), \ell'_n(x; \xi) \right] &= \mathbb{E}_\xi \left[\left(\widehat{\xi}(x) - \bar{\xi}(\xi) \right) \ell'_n(x; \xi) \right] \\ &= \frac{\partial}{\partial \xi} \int_{\mathcal{X}} \widehat{\xi}(x) p_n(x; \xi) \, dx - \bar{\xi}(\xi) \mathbb{E}_\xi \left[\ell'_n(x; \xi) \right] \\ &= \frac{\partial}{\partial \xi} \int_{\mathcal{X}} \widehat{\xi}(x) p_n(x; \xi) \, dx \\ &= \bar{\xi}'(\xi). \end{aligned}$$

Segundo a desigualdade de Schwarz (DEGROOT; SCHERVISH, 2011), sabemos que

$$\left(\text{Cov}_\xi \left[\widehat{\xi}(x), \ell'_n(x; \xi) \right] \right)^2 \leq \text{Var}_\xi \left[\widehat{\xi}(x) \right] \text{Var}_\xi \left[\ell'_n(x; \xi) \right].$$

Consequentemente, temos que

$$\text{Var}_\xi \left[\widehat{\xi} \right] \geq \frac{\left(\bar{\xi}'(\xi) \right)^2}{nG_n(\xi)},$$

que é conhecida como desigualdade de Cramér-Rao. Para um estimador não enviesado, é verdade que $\bar{\xi}(\xi) = \xi$, de maneira que $\bar{\xi}'(\xi) = 1$ e a desigualdade acima se escreve como

$$\text{Var}_\xi \left[\widehat{\xi} \right] \geq \frac{1}{nG_n(\xi)}.$$

A variância de um operador é uma medida de quão dispersas as estimativas estão ao redor da média dos valores obtidos, que no caso de um estimador não enviesado, coincide com o valor verdadeiro do parâmetro. Assim, a partir da construção feita acima, podemos perceber que quanto maior for $G(\xi)$, melhor será a estimativa feita, de maneira que é coerente chamar $G(\xi)$ de informação ou, mais precisamente, de informação de Fisher, uma vez que ela, de fato, quantifica informação, essa última tomada no sentido intuitivo. Além disso, vemos que a definição de G pode ser entendida partindo de um contexto empírico, tendo um significado que se estende para além da geometria riemanniana.

2.2.2 A informação e a métrica de Fisher

Motivados a chamar 2.3 de informação de Fisher devido ao seu significado no contexto de metrologia, definamos-na, agora, com mais precisão.

Definição 3 (Matriz de informação de Fisher). (AMARI; NAGAOKA, 2000) Seja $S = \{p_\xi\}$ um modelo estatístico n -dimensional. Chamamos de matriz de informação de Fisher de S em ξ , denotada por $G(\xi)$, a matriz $n \times n$ cujas entradas $G_{ij}(\xi)$, $i, j \in J$, são dadas por

$$G_{ij}(\xi) := \int_{\mathcal{X}} \partial_i \ell(x; \xi) \partial_j \ell(x; \xi) p(x; \xi) \, dx, \quad (2.4)$$

onde $\ell_\xi(x) := \ell(x; \xi) = \ln p(x; \xi)$ e, quando $n = 1$, $G(\xi)$ é denominada de informação de Fisher de S em ξ .

Perceba que $G(\xi)$ pode divergir se assumirmos $\partial_i \ell_\xi \in L^1(\mathcal{X})$, $i \in I$. Para lidar com esse inconveniente, costuma-se assumir $\partial_i \ell_\xi \in L^2(\mathcal{X})$, $i \in J$, o que se justifica pela desigualdade Cauchy-Schwarz. Além disso, $G(\xi)$, aqui, corresponde ao mesmo $G(\xi)$ de 2.2.1 para o caso de apenas uma observação. Ademais, como foi visto anteriormente, essa expressão pode ser reescrita através do cálculo da média, tal qual foi feito em 2.3. Dessa forma, 2.4 pode ser expressa como

$$G_{ij}(\xi) = E_\xi [\partial_i \ell_\xi \partial_j \ell_\xi].$$

A primeira suposição que foi feita no capítulo anterior permite que a integrabilidade do produto $\partial_i \ell_\xi \partial_j \ell_\xi p_\xi$ seja substituída, de forma equivalente, pela integrabilidade de $\partial_i \partial_j \ell_\xi p_\xi$, com $i, j \in J$, de maneira que 2.4 pode ser reescrita como

$$G_{ij}(\xi) = -E_\xi [\partial_i \partial_j \ell_\xi]. \tag{2.5}$$

De fato,

$$\begin{aligned} E_\xi [\partial_i \ell_\xi \partial_j \ell_\xi] &= \int_{\mathcal{X}} \frac{\partial_i p(x; \xi)}{p(x; \xi)} \frac{\partial_j p(x; \xi)}{p(x; \xi)} p(x; \xi) dx - \underbrace{\int_{\mathcal{X}} \partial_i \partial_j p(x; \xi) dx}_{\partial_j E_\xi [\partial_i \ell_\xi] = 0} \\ &= - \int_{\mathcal{X}} \partial_i \left(\frac{\partial_j p(x; \xi)}{p(x; \xi)} \right) p(x; \xi) dx \\ &= -E_\xi [\partial_i \partial_j \ell_\xi]. \end{aligned}$$

Já a suposição feita sobre o suporte das distribuições de S garante que o denominador de

$$\partial_i \ell(x; \xi) = \frac{\partial_i p(x; \xi)}{p(x; \xi)}$$

não se anule, para $i \in J$.

Supondo, agora, a independência linear do conjunto¹² $\{\partial_1, \dots, \partial_n\}$ para todo $\xi \in \Xi$, então $G(\xi)$ é positivo-definida em Ξ . Consequentemente, como G é simétrica e a função $\xi \mapsto G(\xi)$ é suave, G define unicamente uma métrica riemanniana $\langle \cdot, \cdot \rangle$ em S , também denotada por g , chamada de métrica de Fisher¹³.

Definição 4 (Métrica de Fisher). (AMARI; NAGAOKA, 2000) Seja S um modelo estatístico n -dimensional. Chamamos de métrica de Fisher a métrica riemanniana dada por

$$g(\xi)(X, Y) = E_\xi \left[\left(\frac{\partial \ell_\xi}{\partial X} \right) \left(\frac{\partial \ell_\xi}{\partial Y} \right) \right] \tag{2.6}$$

para todo $\xi \in \Xi$ e para todos os vetores X, Y pertencentes ao espaço tangente $T_{\phi(\xi)}S$ de S em $\phi(\xi)$. Na base $\{\partial_1, \dots, \partial_n\}$, tal métrica é dada por $G(\xi)$.

¹² Essa é uma condição de não singularidade do modelo estatístico visto como variedade diferenciável de dimensão n .

¹³ Essa métrica pode aparecer na literatura, também, como informação métrica.

Como vimos na motivação 2.2.1, a informação de Fisher não se estabelece como uma definição artificial. Agora, com o intuito de fornecer um exemplo de como a matriz de informação de Fisher, em si, aparece espontaneamente em determinados contexto, vamos partir da entropia relativa

$$D_{KL}(p_{\xi_1}; p_{\xi_2}) := \int_{\mathcal{X}} p_{\xi_1}(x) \ln \frac{p_{\xi_1}(x)}{p_{\xi_2}(x)} dx$$

entre duas distribuições p_{ξ_1} , p_{ξ_2} de um modelo estatístico¹⁴ S (NIELSEN, 2022). Ela também é chamada de divergência de Kullback–Leibler e, por mais que ela não seja uma métrica, ela possui todas as propriedades de uma quase-métrica em S , exceto a desigualdade triangular, podendo ser interpretada como uma medida de quão diferentes são as distribuições p_{ξ_1} e p_{ξ_2} (MANNING; SCHÜTZE, 1999).

Ao fazermos a expansão em série de Taylor da entropia relativa ao redor do parâmetro ξ , é possível observarmos o aparecimento da matriz de informação de Fisher no seu segundo termo. Com efeito, sendo S um modelo estatístico n -dimensional e considerando $p_{\xi} \in S$, podemos escrever

$$D_{KL}(p_{\xi}; p_{\xi+d\xi}) = \int_{\mathcal{X}} p_{\xi}(x) \ln \frac{p_{\xi}(x)}{p_{\xi+d\xi}(x)} dx.$$

Sabemos que o termo quadrático da expansão em série de Taylor de D_{KL} em torno de ξ é $\frac{1}{2}d\xi^T \text{Hess}(D_{KL})d\xi$, onde $\text{Hess}(D_{KL})$ é a matriz hessiana de D_{KL} em ξ (ARORA, 2011). Devido a isso, calculando as derivadas de segunda ordem da divergência de Kullback-Leibler, obtemos

$$\partial_i \partial_j D_{KL}(p_{\xi}, p_{\xi+d\xi}) = - \int_{\mathcal{X}} p_{\xi}(x) \left[\frac{(\partial_i \partial_j p_{\xi+d\xi})(x)}{p_{\xi+d\xi}(x)} - \frac{(\partial_i p_{\xi+d\xi})(x) (\partial_j p_{\xi+d\xi})(x)}{p_{\xi+d\xi}^2(x)} \right] dx,$$

para todo $i \in J$. Então, no limite em que $d\xi \rightarrow 0$, temos que

$$H_{ij} = \int_{\mathcal{X}} \frac{\partial_i p_{\xi}(x) \partial_j p_{\xi}(x)}{p_{\xi}(x)} dx = G_{ij}(\xi).$$

Dessa maneira, observamos que a matriz de informação de Fisher surge espontaneamente quando ocorrem aproximações locais da entropia relativa (NIELSEN, 2022).

Já a naturalidade¹⁵ da métrica de Fisher pode ser entendida através do Teorema de Caracterização de Chentsov, que garante que ela é a única¹⁶ métrica riemanniana em modelos estatístico sobre conjuntos finitos que é invariante sob morfismos de Markov, ou seja e grosso

¹⁴ Pelo nome, é esperado que D_{KL} relacione-se com a entropia de Shannon e isso, de fato, ocorre. A entropia de Shannon de $p_{\xi_1} \in S$, entendida como uma medida da incerteza média de uma variável aleatória com distribuição p_{ξ_1} , é definida como $h(p_{\xi_1}) := - \int_{\mathcal{X}} p_{\xi_1} \ln p_{\xi_1} dx$ (COVER; THOMAS, 2006; NIELSEN, 2022). Assim, dadas $p_{\xi_1}, p_{\xi_2} \in S$, podemos escrever $D_{KL}(p_{\xi_1}; p_{\xi_2}) = h^{\times}(p_{\xi_1}; p_{\xi_2}) - h(p_{\xi_1})$, onde $h^{\times}(p_{\xi_1}; p_{\xi_2}) := - \int_{\mathcal{X}} p_{\xi_1} \ln p_{\xi_2} dx$ é chamada de entropia cruzada entre p_{ξ_1} e p_{ξ_2} (NIELSEN, 2022).

¹⁵ Durante esse trabalho, quando o termo “naturalidade” surgir, significará que o objeto por ele adjetivo se destaca dentro do contexto estudado.

¹⁶ Mais precisamente, é a única a menos de constantes multiplicativas reais e positivas.

modo, mapas que associam elementos de \mathcal{X} com medidas de probabilidade em um outro outro conjunto \mathcal{X}' (NIELSEN, 2022; AY *et al.*, 2017). Esse teorema será apresentado com mais cuidado em 3.2.

2.2.3 O tensor de Amari-Chentsov

Como citado em 1.1, os trabalhos de Chentsov estudando famílias de distribuições de probabilidade munidas da métrica de Fisher e de uma conexão diferente da conexão de Levi-Civita dela conduziram-no à descoberta de um campo tensorial que, futuramente, viria a desempenhar uma função significativa dentro da geometria da informação. Em determinados contextos, como aqueles advindos dos trabalhos de Lauritzen (LAURITZEN, 1987), esse campo aparece sendo referido como tensor de assimetria, tensor de “skewness” ou tensor 3-simétrico, enquanto que, em referências que tratam de geometria da informação finita, ele pode surgir com o nome de tensor de Amari-Chentsov ou apenas tensor de Amari. Neste trabalhos, optamos por nos referirmos a esse campo tensorial como tensor de Amari-Chentsov quando estivermos tratando de modelos estatísticos, especialmente, em contexto de amostras finitas, e de tensor 3-simétrico quando a abordagem for vertida e estivermos estudando, de maneira mais geral, variedades estatísticas.

Dito isso, sendo uma métrica riemanniana, podemos interpretar a métrica de Fisher como um campo tensorial covariante de *rank* 2 simétrico. Partindo disso, é possível compreender o tensor de Amari-Chentsov como uma versão de *rank* 3 dessa métrica, sendo caracterizado por invariâncias tal qual sua versão 2-covariante. Estabelecida, então, a ideia que conduz a sua definição, apresentemos tal campo.

Definição 5 (Tensor de Amari-Chentsov). (AY *et al.*, 2017) Seja S modelo estatístico. Chamamos de tensor de Amari-Chentsov o campo tensorial 3-covariante simétrico em todas as suas entradas e dado por

$$T(\xi)(X, Y, Z) := \mathbb{E}_{\xi} \left[\left(\frac{\partial \ell_{\xi}}{\partial X} \right) \left(\frac{\partial \ell_{\xi}}{\partial Y} \right) \left(\frac{\partial \ell_{\xi}}{\partial Z} \right) \right] \quad (2.7)$$

para todo $\xi \in \Xi$ e para todos os vetores X, Y, Z pertencentes ao espaço tangente $T_{\phi(\xi)}S$ de S em $\phi(\xi)$.

Dessa forma, em um sistema de coordenadas ϕ , T escreve-se, para todo $\xi \in \Xi$, como

$$T_{ijk}(\xi) = \int_{\mathcal{X}} \partial_i \ell(x; \xi) \partial_j \ell(x; \xi) \partial_k \ell(x; \xi) p(x; \xi) dx.$$

Além disso, note que a convergência da integral de $T(\xi)$ está condicionada às derivadas no argumento da média 2.7 pertencerem a $L^3(\mathcal{X})$, sendo que a justificativa disso segue do mesmo comentário feito sobre integrabilidade quando apresentamos a métrica de Fisher.

2.3 As α -conexões e a relação entre distribuições gaussianas e a geometria hiperbólica

Como foi destacado no início em 1.1, nos precedentes da descoberta do tensor de Amari-Chentsov, Chentsov estudou os modelos estatísticos rompendo com o paradigma que estabelecia a conexão de Levi-Civita como a mais adequada para esse estudo. Posteriormente, Amari e Nagaoka generalizariam essa abordagem vanguardista por meio da introdução das α -conexões: uma família uniparamétrica de conexões lineares que engendrava tanto as descobertas de Efron e Dawid quanto a conexão de Levi-Civita. Sendo, então, uma generalização dessa última e, devido ao fato das conexões lineares serem determinadas por seus símbolos de Christoffel, é coerente calcular os símbolos da conexão de Levi-Civita da métrica de Fisher para, assim, visualizar como as α -conexões irão a englobar. Para tal, é fácil ver que

$$\partial_i g_{jk} = E_{\xi} [(\partial_i \partial_j \ell_{\xi}) (\partial_k \ell_{\xi})] + E_{\xi} [(\partial_j \ell_{\xi}) (\partial_i \partial_k \ell_{\xi})] + E_{\xi} [(\partial_i \ell_{\xi}) (\partial_j \ell_{\xi}) (\partial_k \ell_{\xi})].$$

Por simples mudança de índices, essa expressão nos fornece como $\partial_j g_{ik}$ e $\partial_k g_{ij}$ escrevem-se em coordenadas. Assim,

$$\begin{aligned} \Gamma_{ijk}(\xi) &= \sum_{m \in I} \Gamma_{ij}^m(\xi) g_{mk}(\xi) \\ &= \frac{1}{2} (\partial_i g_{ik}(\xi) + \partial_j g_{ik}(\xi) - \partial_k g_{ij}(\xi)) \\ &= E_{\xi} [(\partial_i \partial_j \ell_{\xi}) (\partial_k \ell_{\xi})] + \frac{1}{2} E_{\xi} [(\partial_i \ell_{\xi}) (\partial_j \ell_{\xi}) (\partial_k \ell_{\xi})] \end{aligned}$$

Perceba que uma perturbação no segundo termo acima, digamos por $-\alpha/2$, com $\alpha \in [-1, 1]$, ainda produz uma conexão. Essa é ideia subjacente às α -conexões.

Definição 6 (α -conexão). (AY *et al.*, 2017) Seja S um modelo estatístico e $\alpha \in [-1, 1]$. As α -conexões $\nabla^{(\alpha)}$ são conexões lineares em S determinadas por

$$\Gamma_{ijk}(\xi) = E_{\xi} [(\partial_i \partial_j \ell_{\xi}) (\partial_k \ell_{\xi})] + \frac{1-\alpha}{2} E_{\xi} [(\partial_i \ell_{\xi}) (\partial_j \ell_{\xi}) (\partial_k \ell_{\xi})] \quad (2.8)$$

para todo $\xi \in \Xi$. Elas compõem uma família de conexões parametrizadas por α .

Note que $\nabla^{(0)}$ coincide com a conexão de Levi-Civita, como esperado. É importante ressaltar que essas conexões serão melhor desenvolvidas nos próximos dois capítulos, onde elas serão rerepresentadas através de duas perspectivas distintas: no capítulo 3, em um contexto de amostras finitas e, no capítulo 4, através da noção de dualidade entre conexões, onde comentaremos um pouco mais sobre a restrição de α a $[-1, 1]$. Entretanto, uma primeira característica evidente dessa família de conexões é a ausência de torção por parte de seus elementos. Da geometria riemanniana, sabemos que uma conexão linear é livre de torção se seus símbolos de Christoffel Γ_{ij}^k são simétricos em i e j em algum sistema de coordenadas, para todo $i, j \in I$ (AY

et al., 2017). Observando 2.8, vemos que isso é satisfeito para as α -conexões devido à simetria de derivadas segundas de ℓ_ξ advinda do Teorema de Schwarz.

Ademais, as α -conexões podem ser interpretadas a partir dos desenvolvimentos de Jeffreys a respeito de aproximações locais da divergência de Kullback-Leibler. Em seus estudos, o termo quadrático em que aparecia a matriz de informação de Fisher foi interpretado como o elemento quadrático de comprimento riemanniano ds_ξ induzido por essa métrica (NIELSEN, 2022). Em outras palavras, da expansão em série de Taylor de D_{KL} ao redor de $p_\xi \in S$, ele obteve

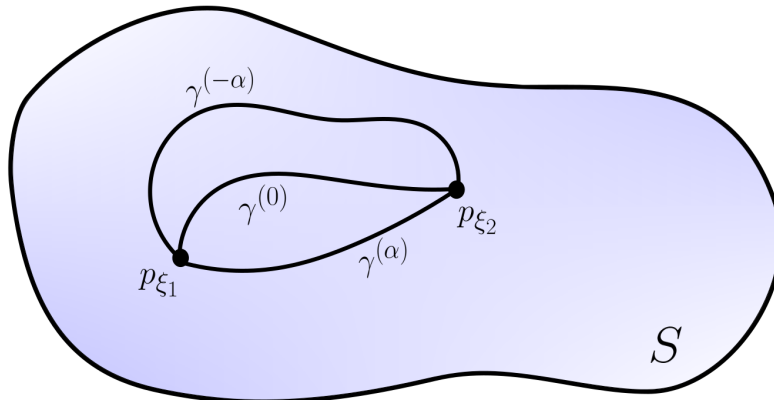
$$ds_\xi = \sqrt{d\xi^T G d\xi}. \tag{2.9}$$

Desse modo, *a priori*, a maneira mais natural de se medir dissimilaridade entre duas distribuições p_{ξ_1} e p_{ξ_2} de um modelo estatístico S é através da distância de Fisher-Rao ρ_F , que consiste naquela obtida via ds por

$$\rho_F(p_{\xi_1}, p_{\xi_2}) := \inf_{\gamma} \int_0^1 ds_{\gamma(t)} dt,$$

onde $\gamma(t)$ é geodésica riemanniana com as condições de contorno $\gamma(0) = p_{\xi_1}$ e $\gamma(1) = p_{\xi_2}$ (NIELSEN, 2022). Assim, as α -conexões permitem mensurar a dissimilaridade entre distribuições de um modelo, perturbando a noção de distância riemanniana com o intuito de que ela se adapte melhor ao modelo que está sob análise. Como exemplo disso, temos a família exponencial, que consiste em uma família de distribuições de probabilidade cuja conexão usualmente considerada apropriada para lidar com ela é a conexão exponencial $\nabla^{(1)}$ devido à sua propriedade de ser plana nesse modelo. Outro exemplo é a conexão mistura $\nabla^{(-1)}$, que apresenta essa mesma característica na família mistura, uma outra família de distribuições de probabilidade.

Figura 2 – Representação gráfica de diferentes medidas de dissimilaridade entre duas distribuições p_{ξ_1} e p_{ξ_2} de um modelo estatístico S do ponto de vista da α -geometria. Supondo suas existências, $\gamma^{(0)}$, $\gamma^{(-\alpha)}$ e $\gamma^{(\alpha)}$ representam, respectivamente, geodésicas relativas a $\nabla^{(0)}$, $\nabla^{(-\alpha)}$ e $\nabla^{(\alpha)}$ ligando essas distribuições



Fonte: Adaptada de Nielsen (2022, p 4).

Observe que 2.9 é invariante sob reparametrizações suaves invertíveis (NIELSEN, 2022). De fato, denotemos por G^ξ a matriz da métrica de Fisher no sistema de coordenadas ξ . Considerando um segundo sistema η , temos que essa matriz, escrita nele e denotada por G^η , satisfaz

$$G^\xi = \left[\frac{\partial \eta_i}{\partial \xi_j} \right]_{ij}^T G^\eta \left[\frac{\partial \eta_i}{\partial \xi_j} \right]_{ij}.$$

Sendo a mudança entre ξ e η suave, sabemos que, para todo $i \in J$,

$$d\eta_i = \sum_{j \in J} \frac{\partial \eta_i}{\partial \xi_j} d\xi_j.$$

Assim, o termo interior à raiz em 2.9 pode ser expresso como

$$d\xi^T \left[\frac{\partial \eta_i}{\partial \xi_j} \right]_{ji} G^\eta \left[\frac{\partial \eta_i}{\partial \xi_j} \right]_{ij} d\xi = d\eta^T G^\eta d\eta,$$

e, conseqüentemente, $ds_\xi = ds_\eta$. Por conseguinte, no caso da família de distribuições gaussianas, parametrizações pela média μ junto à variância σ^2 , ao desvio padrão σ ou a $\sqrt{2}\sigma$, produzem a mesma distância de Rao. Dito isso, estudemos, brevemente, qual estrutura geométrica emerge do modelo das distribuições gaussianas quando partimos da parametrização $(\mu, \sqrt{2}\sigma)$. Considere, então,

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

com $\mu \in \mathbb{R}$, $\sigma \in (0, \infty)$. Assim,

$$g_{\mu\mu} = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \mu} \ln \left(\frac{1}{\sqrt{\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} \right) \right)^2 \frac{1}{\sqrt{\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} dx = \frac{2}{\sigma^2}$$

$$g_{\sigma\sigma} = \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \sigma} \ln \left(\frac{1}{\sqrt{\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} \right) \right)^2 \frac{1}{\sqrt{\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} dx = \frac{2}{\sigma^2}$$

$$\begin{aligned} g_{\mu\sigma} &= g_{\sigma\mu} \\ &= \int_{-\infty}^{\infty} \left(\frac{\partial}{\partial \sigma} \ln \left(\frac{1}{\sqrt{\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} \right) \right) \left(\frac{\partial}{\partial \mu} \ln \left(\frac{1}{\sqrt{\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} \right) \right) \frac{1}{\sqrt{\pi}\sigma} e^{-\frac{(x-\mu)^2}{\sigma^2}} dx \\ &= 0 \end{aligned}$$

Logo, a matriz de informação de Fisher $G(\mu, \sigma) := G((\mu, \sigma))$, nesse caso, escreve-se como

$$G(\mu, \sigma) = \frac{2}{\sigma^2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

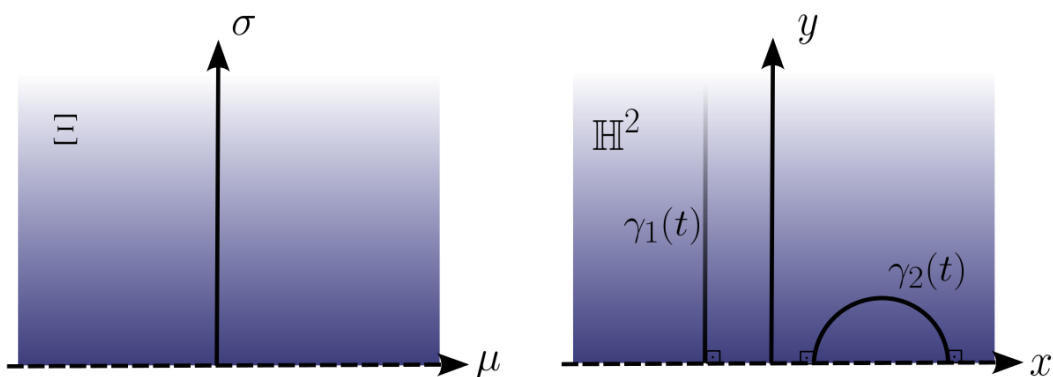
Essa é uma expressão célebre, associada a uma estrutura geométrica reconhecida desde o final do século XIX, sendo essa a estrutura hiperbólica. Ela pode ser apresentada, sumariamente,

por meio da história do Postulado das Paralelas. Sendo um dos postulados introduzidos por Euclides na clássica obra “Os elementos” (EUCLID, 2008), ele teve atrelado a si uma sequência de tentativas falhas de o demonstrar, como é possível observar nos trabalhos do britânico John Playfair, do francês Adrien-Marie Legendre e do húngaro Farkas Bolyai (LEWIS, 1920). Entretanto, entre as décadas de 1820 e 1830, o filho de Farkas, János Bolyai, partindo da negação do postulado, desenvolveu uma geometria completamente compatível com os demais quatro postulados da axiomática de Euclides, como suspeitava o próprio Gauss (LEWIS, 1920). Em complemento a isso, é importante pontuar que a elaboração do cerne do que hoje é conhecido como geometria hiperbólica não está restrito a János, mas se encontra presente nos trabalhos de seu contemporâneo russo, Nikolai Lobachevsky (DANI; PAPADOPOULOS, 2019), que, inclusive, nomeia um dos modelos bidimensionais para essa geometria. O Plano de Lobachevsky, também conhecido como Semiplano de Poincaré, consiste no semiplano superior $\mathbb{H}^2 := \{(x, y) \in \mathbb{R}^2 \mid y > 0\}$ munido da estrutura de variedade herdade de \mathbb{R}^2 e da métrica riemanniana expressa em coordenadas retangulares como

$$g_{\mathbb{H}^2}(x, y) = \frac{1}{y^2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Perceba que \mathbb{H}^2 e o espaço de parâmetros Ξ das distribuições gaussianas são difeomorfos. Observando que $G(\mu, \sigma) = 2g_{\mathbb{H}^2}(\mu, \sigma)$, concluímos que, a menos de um fator conforme, a métrica de Fisher no modelo estatístico em questão coincide com a métrica hiperbólica do modelo do Plano de Lobachevsky. Desse modo, informações sobre a geometria hiperbólica nos fornecem conhecimento sobre a geometria presente no modelo gaussiano. Por exemplo, sabemos que a curvatura desse último, visto como variedade riemanniana, é constante e estritamente negativa. Além disso, as geodésicas respectivas à conexão de Levi-Civita nesse espaço são imagens, pelo sistema de coordenadas (a menos da reparametrização), de semirretas ou semicírculos perpendiculares a $\sigma = 0$.

Figura 3 – Representação gráfica da semelhança entre \mathbb{H}^2 e Ξ no caso da família de distribuições gaussianas. Em \mathbb{H}^2 , estão ilustradas as curvas γ_1 e γ_2 , que contemplam o formato das geodésicas de $(\mathbb{H}^2, g_{\mathbb{H}^2})$



Fonte: Elaborada pelo autor.

Por conseguinte, os comentários acima expõem uma relação direta entre dois objetos que são importantes em áreas consideradas distintas: um na estatística e outro na geometria riemanniana. Por um lado, através de [1](#), as distribuições gaussianas podem ser interpretadas como naturais no tratamento de grandes amostras, quando se há a hipótese de independência e igual distribuição. Por outro lado, a geometria hiperbólica se destaca como uma das três geometrias que, em dimensão dois, caracterizam superfícies conexas fechadas através de curvatura constante.

GEOMETRIA DA INFORMAÇÃO FINITA

3.1 Introdução a uma abordagem extrínseca: apresentação de $\mathcal{P}_+(\mathcal{X})$ através de álgebra de funções sobre um conjunto finito

A geometria da informação apresentada em 2 possui um caráter geométrico intrínseco de análise dos modelos estatísticos. Em outras palavras, os modelos estatísticos são compreendidos como variedades diferenciáveis e neles são definidos objetos de caráter geométrico, como a métrica de Fisher, o tensor de Amari-Chentsov e as α -conexões. Contudo, abordagens mais recentes partem de uma perspectiva extrínseca de análise, em que os modelos estatísticos perdem parte de seu protagonismo como objetos estatísticos e são tomados já inicialmente como variedades. Isso significa que os espaços de parâmetros são mergulhados em espaços de medidas e a geometria deles é definida em termos desses mergulhos. Nos casos em que os espaços de medidas possuem dimensão infinita, surgem diversas dificuldades técnicas em que, usualmente, são utilizadas variedades de Banach para as resolver (AY *et al.*, 2017). Entretanto, no caso finito, elementos de análise nessas variedades são dispensáveis. Frente a isso, dados os ganhos teóricos de contemplar uma abordagem extrínseca, apresentaremos-na neste capítulo no caso finito. É importante pontuar que, por mais próximas que sejam algumas construções, resultados e definições no caso intrínseco e no extrínseco, como destaca (AY *et al.*, 2017), essas abordagens não são completamente equivalentes.

Dito isso, seja $\mathcal{X} = \{x_1, x_2, \dots, x_{n+1}\}$ um conjunto finito não vazio. Seja, também, $\mathcal{F}(\mathcal{X})$ a álgebra de funções de \mathcal{X} a \mathbb{R} , cuja base canônica é denotada por $\mathcal{B}_{\mathcal{F}(\mathcal{X})} = \{e_1, \dots, e_{n+1}\}$. Por base canônica, entende-se as funções $e_i : \mathcal{X} \rightarrow \mathbb{R}$ tais que $e_i(x_j) = \delta_{ij}$, onde δ_{ij} representa a delta de Kronecker de i e j . Note que $\mathcal{F}(\mathcal{X})$ é isomorfo a \mathbb{R}^{n+1} . Sendo $\mathbb{1}$ a identidade dessa álgebra, usaremos a notação \mathbb{R} para nos referirmos ao

subespaço vetorial $\mathbb{R} \cdot \mathbf{1}$ de $\mathcal{F}(\mathcal{X})$ composto pelas funções constantes, ou seja,

$$\mathbb{R} := \mathbb{R} \cdot \mathbf{1} := \{c \cdot \mathbf{1} \in \mathcal{F}(\mathcal{X}) \mid c \in \mathbb{R}\}.$$

Isso posto, denotando por¹ $\mathcal{S}(\mathcal{X}) = \mathcal{F}(\mathcal{X})^*$ o espaço dual de $\mathcal{F}(\mathcal{X})$, podemos interpretar as formas lineares $a \in \mathcal{S}(\mathcal{X})$ como medidas assinadas² sobre a σ -álgebra discreta $\mathcal{P}(\mathcal{X})$ de \mathcal{X} . Isso se deve ao fato presente em (YEH, 2014) e que consiste em, dado $a \in \mathcal{S}(\mathcal{X})$, existir uma única medida de Radon $\lambda : \mathcal{P}(\mathcal{X}) \rightarrow \mathbb{R}$ assinada e finita tal que, para toda $f \in \mathcal{F}(\mathcal{X})$,

$$a(f) = \sum_{i \in I} f(x_i) \lambda(x_i).$$

Isso pode ser pensado como uma consequência do Teorema da Representação de Riesz e não significa que os elementos de $\mathcal{S}(\mathcal{X})$ são medidas, mas que, escritos em $B_{\mathcal{S}(\mathcal{X})} = \{\delta^1, \dots, \delta^{n+1}\}$ base dual de $B_{\mathcal{F}(\mathcal{X})}$, o i -ésimo coeficiente a_i de a é igual a $\lambda(x_i)$.

Colocado isso, é possível definir algumas variedades úteis, que consistem na variedade de medidas finitas positivas sobre³ \mathcal{X} , denotada por $\mathcal{M}_+(\mathcal{X})$, e na variedade de medidas de probabilidade positivas sobre \mathcal{X} , conhecida por $\mathcal{P}_+(\mathcal{X})$. Mais precisamente, considerando os elementos de $\mathcal{S}(\mathcal{X})$ escritos em $B_{\mathcal{S}(\mathcal{X})}$,

$$\begin{aligned} \mathcal{M}_+(\mathcal{X}) &:= \left\{ \mu \in \mathcal{S}(\mathcal{X}) \mid \mu_i > 0, i \in I \right\}, \\ \mathcal{P}_+(\mathcal{X}) &:= \left\{ \xi \in \mathcal{M}_+(\mathcal{X}) \mid \sum_{i \in I} \xi_i = 1 \right\}. \end{aligned}$$

Como $\mathcal{M}_+(\mathcal{X})$ é subvariedade aberta de $\mathcal{S}(\mathcal{X})$, a diferencial da função inclusão em cada ponto de $\mathcal{M}_+(\mathcal{X})$ é isomorfismo e, conseqüentemente, os fibrados tangente e cotangente dela são

$$T\mathcal{M}_+(\mathcal{X}) = \mathcal{M}_+(\mathcal{X}) \times \mathcal{S}(\mathcal{X}), \quad T^*\mathcal{M}_+(\mathcal{X}) = \mathcal{M}_+(\mathcal{X}) \times \mathcal{F}(\mathcal{X}),$$

uma vez que sabemos que, como $\mathcal{S}(\mathcal{X})$ é espaço vetorial,

$$T\mathcal{S}(\mathcal{X}) = \mathcal{S}(\mathcal{X}) \times \mathcal{S}(\mathcal{X}), \quad T^*\mathcal{S}(\mathcal{X}) = \mathcal{S}(\mathcal{X}) \times \mathcal{F}(\mathcal{X}).$$

Calculemos, então, o fibrado tangente de $\mathcal{P}_+(\mathcal{X})$. Para tal, consideremos a carta global $(\varphi^{-1}(\Xi), \varphi)$ de $\mathcal{P}_+(\mathcal{X})$ dada por

$$\varphi : \mathcal{P}_+(\mathcal{X}) \rightarrow \Xi, \quad \xi = \sum_{i \in I} \xi_i \delta^i \mapsto (\xi_1, \dots, \xi_n), \quad (3.1)$$

¹ Por vezes, escreveremos $\mathcal{S}(\mathcal{X})^* = \mathcal{F}(\mathcal{X})$ ou $\mathcal{S}(\mathcal{X}) = \mathcal{F}(\mathcal{X})^*$, onde as igualdades devem ser entendidas a menos de isomorfismos.

² Alguns autores se referem a medidas assinadas sobre uma σ -álgebra como cargas, em diálogo com a física (BARTLE, 1995).

³ Por vezes, na literatura, $\mathcal{M}_+(\mathcal{X})$ pode ser referida como cone positivo (CAMPBELL, 1986).

onde Ξ é o aberto de \mathbb{R}^n dado por

$$\Xi = \left\{ (\xi_1, \dots, \xi_n) \in \mathbb{R}^n \mid \xi_i > 0, i = 1, \dots, n, \sum_{i \in J} \xi_i < 1 \right\}. \quad (3.2)$$

Sejam, então, $\xi \in \mathcal{P}_+(\mathcal{X})$ e $c : (-\varepsilon, \varepsilon) \rightarrow \mathcal{P}_+(\mathcal{X})$ curva suave tal que $c(0) = \xi$. Escrevendo

$$c(t) = \sum_{i \in I} c_i(t) \delta^i,$$

temos que

$$\sum_{i \in I} c'_i(t) = \frac{d}{dt} \underbrace{\sum_{i \in I} c_i(t)}_1 = 0.$$

Desse modo, obtemos que

$$T_\xi \mathcal{P}_+(\mathcal{X}) = \{\xi\} \times \mathcal{S}_0(\mathcal{X}),$$

onde

$$\mathcal{S}_0(\mathcal{X}) := \left\{ \sum_{i \in I} \xi_i \delta^i \in \mathcal{S}(\mathcal{X}) \mid \sum_{i \in I} \xi_i = 0 \right\}$$

e, conseqüentemente,

$$T \mathcal{P}_+(\mathcal{X}) = \mathcal{P}_+(\mathcal{X}) \times \mathcal{S}_0(\mathcal{X}).$$

A partir disso, calculemos o fibrado cotangente de $\mathcal{P}_+(\mathcal{X})$. Obviamente,

$$T^* \mathcal{P}_+(\mathcal{X}) = \mathcal{P}_+(\mathcal{X}) \times \mathcal{S}_0(\mathcal{X})^*,$$

contudo, vamos escrever $\mathcal{S}_0(\mathcal{X})$ em termos de $\mathcal{F}(\mathcal{X})$, já que $\mathcal{S}_0(\mathcal{X}) \subset \mathcal{S}(\mathcal{X}) = \mathcal{F}(\mathcal{X})^*$.

Para tal, defina a função

$$\begin{aligned} \rho : \mathcal{F}(\mathcal{X}) &\longrightarrow \mathcal{S}_0(\mathcal{X})^* \\ f &\longmapsto \psi, \end{aligned}$$

onde ψ é o funcional

$$\begin{aligned} \psi : \mathcal{S}_0(\mathcal{X}) &\longrightarrow \mathbb{R} \\ a &\longmapsto a(f). \end{aligned}$$

Podemos notar que o núcleo de ρ são as funções constantes, ou seja, $\ker \rho = \mathbb{R}$. Para tal, perceba que $\dim \mathcal{S}_0(\mathcal{X})^* = \dim \mathcal{P}_+(\mathcal{X}) = n$. Assim, pelo Teorema do Núcleo e da Imagem aplicado a ρ , temos que $\ker \rho$ é o subespaço vetorial de $\mathcal{F}(\mathcal{X})$ de dimensão unitária, isto é, o subespaço gerado por $\mathbb{1}$. Dessa maneira, fica evidente que $\ker \rho = \mathbb{R}$. Aplicando o Primeiro Teorema do Isomorfismo⁴, temos que $\mathcal{S}_0(\mathcal{X})^*$ é isomorfo a $\mathcal{F}(\mathcal{X})/\mathbb{R}$, onde podemos destacar o isomorfismo natural

$$\begin{aligned} \bar{\rho} : \mathcal{F}(\mathcal{X})/\mathbb{R} &\longrightarrow \mathcal{S}_0(\mathcal{X})^* \\ f + \mathbb{R} &\longmapsto \rho(f). \end{aligned}$$

⁴ Por Primeiro Teorema do Isomorfismo, entende-se o teorema que afirma que, dada $\rho : V \rightarrow W$ uma transformação linear entre os espaços vetoriais V e W , então $V/\ker(\rho)$ é isomorfo a $\text{Im}(\rho)$. (ROMAN, 2007)

Logo, concluímos que

$$T^* \mathcal{P}_+(\mathcal{X}) = \mathcal{P}_+(\mathcal{X}) \times \mathcal{F}(\mathcal{X})/\mathbb{R}.$$

Por completude da exposição, podemos expressar $\bar{\rho}^{-1}$ em termos de $\mathcal{B}_{\mathcal{S}(\mathcal{X})}$. Para tal, dada $f \in \mathcal{S}_0(\mathcal{X})^*$, consideremos uma extensão⁵ $\tilde{f} \in \mathcal{S}(\mathcal{X})^*$, de maneira que

$$\bar{\rho}^{-1}(f) = \left(\sum_{i \in J} f^i e_i \right) + \mathbb{R},$$

onde $f^i := \tilde{f}(\delta^i)$, $i \in I$. Para mostrar que a expressão acima está bem definida, dada $f \in \mathcal{S}_0(\mathcal{X})^*$, considere \tilde{f}_1 e \tilde{f}_2 extensões de f a $\mathcal{S}(\mathcal{X})$. Como $\mathcal{S}(\mathcal{X}) = \mathcal{S}_0(\mathcal{X}) \oplus \mathcal{S}_0(\mathcal{X})^\perp$, para todo $a \in \mathcal{S}_0(\mathcal{X})$, sabemos que

$$a^\perp = 0 = \sum_{i \in I} a_i (\delta^i)^\perp,$$

onde $(\delta^i)^\perp$ indica a projeção perpendicular de δ^i em $\mathcal{S}_0(\mathcal{X})$. Assim, como $\sum_{i \in I} a_i = 0$, temos que $(\delta^i)^\perp$ são iguais para todo⁶ $i \in I$. Isso posto, invocando a linearidade de \tilde{f}_1 e \tilde{f}_2 , obtemos que

$$\begin{aligned} \sum_{i \in I} \tilde{f}_1(\delta^i) e_i + \mathbb{R} &= \sum_{i \in I} \left[\tilde{f}_1((\delta^i)^\perp) + \tilde{f}_1((\delta^i)^\top) \right] e_i + \mathbb{R} \\ &= \sum_{i \in I} \left[\tilde{f}_1((\delta^i)^\perp) + \tilde{f}_2((\delta^i)^\top) \right] e_i + \mathbb{R} \\ &= \sum_{i \in I} \left[\tilde{f}_2((\delta^i)^\perp) + \tilde{f}_2((\delta^i)^\top) - c \right] e_i + \mathbb{R} \\ &= \sum_{i \in I} \tilde{f}_2(\delta^i) e_i + \mathbb{R}, \end{aligned}$$

onde $c := \tilde{f}_2((\delta^i)^\perp) - \tilde{f}_1((\delta^i)^\perp)$, de maneira que $\tilde{\rho}^{-1}(f)$, como foi apresentada, independe da extensão de f tomada.

Ademais, calculemos agora os vetores coordenados de $\mathcal{P}_+(\mathcal{X})$. Perceba que a inversa de 3.1 é dada por

$$\varphi^{-1} : \mathbb{E} \rightarrow \mathcal{P}_+(\mathcal{X}), \quad (x_1, x_2, \dots, x_n) \mapsto \sum_{i \in J} x_i \delta^i + \left(1 - \sum_{i \in J} x_i \right) \delta^{n+1}.$$

Desse modo, temos que

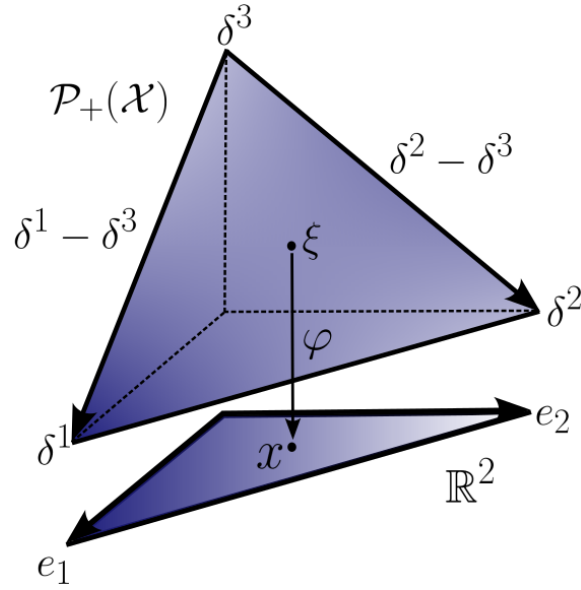
$$\partial_i(\xi) := \frac{\partial}{\partial x_i}(\xi) = \frac{\partial \varphi^{-1}}{\partial x_i}(\varphi(\xi)) = \delta^i - \delta^{n+1}, \quad i \in J,$$

são tais vetores coordenados.

⁵ A existência dessa extensão pode ser vista como consequência do Teorema de Hahn-Banach.

⁶ Os elementos $f \in \ker \rho$ são tais que, para todo $a \in \mathcal{S}_0(\mathcal{X})$, $a(f) = \sum_{i \in I} a_i f^i = 0$. Note que essa última igualdade é análoga a $\sum_{i \in I} a_i (\delta^i)^\perp = 0$. Desse modo, observamos que $(\delta^i)^\perp$ são iguais para todo $i \in I$.

Figura 4 – Representação gráfica de $\mathcal{P}_+(\mathcal{X})$ para $n = 2$. É possível observar o aspecto de 2-simplexo de $\mathcal{P}_+(\mathcal{X})$



Fonte: Adaptada de [Ay et al. \(2017, p 28\)](#).

Além disso, como o conjunto $\{\delta^i - \delta^{n+1} \mid i \in J\}$ é base para $\mathcal{S}_0(\mathcal{X})$, podemos calcular sua base dual, dada pelos elementos

$$dx_i := e_i + \mathbb{R}, \quad i \in J.$$

Desse modo, para todo $f + \mathbb{R}$, temos que

$$\begin{aligned} f + \mathbb{R} &= \left(\sum_{i \in I} f^i e_i \right) + \mathbb{R} \\ &= \sum_{i \in J} f^i (e_i + \mathbb{R}) + f^{n+1} \left[\left(\mathbb{1} - \sum_{i \in J} e_i \right) + \mathbb{R} \right] \\ &= \sum_{i \in J} f^i (e_i + \mathbb{R}) - \sum_{i \in J} f^{n+1} (e_i + \mathbb{R}) \\ &= \sum_{i \in J} (f^i - f^{n+1}) dx_i, \end{aligned}$$

ou seja, nessa base, os elementos $f + \mathbb{R}$ possui coeficientes $f^i - f^{n+1}$ para $f \in \mathcal{F}(\mathcal{X})$ ([AY et al., 2017](#)).

3.2 A métrica de Fisher

Nesta seção, vamos introduzir a métrica de Fisher não apenas em $\mathcal{P}_+(\mathcal{X})$, mas em $\mathcal{M}_+(\mathcal{X})$. Para tal, comecemos estabelecendo, por meio da seguinte definição, o mapa $\tilde{\phi}_\mu$, que nos será útil.

Definição 7. (AY *et al.*, 2017) Dado $\mu \in \mathcal{M}_+(\mathcal{X})$, definimos o produto $\langle \cdot, \cdot \rangle_\mu$ em $\mathcal{F}(\mathcal{X})$ dado por

$$\langle f, g \rangle_\mu = \mu(fg) \in \mathbb{R}, \quad f, g \in \mathcal{F}(\mathcal{X}), \quad (3.3)$$

onde fg corresponde ao produto de funções da álgebra $\mathcal{F}(\mathcal{X})$.

Note que o produto acima assemelha-se a uma métrica riemanniana em $\mathcal{M}_+(\mathcal{X})$, contudo, tomando valores no espaço cotangente de $\mathcal{M}_+(\mathcal{X})$. Assim, podemos pensar em isomorfismos entre $\mathcal{F}(\mathcal{X})$ e $\mathcal{S}(\mathcal{X})$ que desempenham a função dos isomorfismos musicais em um contexto legitimamente riemanniano para 3.3 em $\mathcal{M}_+(\mathcal{X})$. O primeiro consiste no isomorfismo,

$$\mathcal{F}(\mathcal{X}) \longrightarrow \mathcal{S}(\mathcal{X}), \quad f \longmapsto \langle f, \cdot \rangle_\mu,$$

enquanto o segundo, que compreende o inverso desse primeiro, é dado por

$$\tilde{\phi}_\mu : \mathcal{S}(\mathcal{X}) \rightarrow \mathcal{F}(\mathcal{X}), \quad \tilde{\phi}_\mu(a) = \frac{da}{d\mu} := \sum_{i \in I} \frac{a_i}{\mu_i} e_i. \quad (3.4)$$

Isso nos permite definir em $\mathcal{S}(\mathcal{X})$ um produto análogo a 3.3.

Definição 8. (AY *et al.*, 2017) Dado $\mu \in \mathcal{M}_+(\mathcal{X})$, definimos o produto em $\mathcal{S}(\mathcal{X})$ dado por

$$\langle a, b \rangle_\mu := \mu \left(\frac{da}{d\xi} \frac{db}{d\xi} \right) = \sum_{i \in I} \frac{a_i}{\mu_i} \frac{b_i}{\mu_i} \mu_i$$

para todo $a, b \in \mathcal{S}(\mathcal{X})$, onde $\frac{da}{d\mu} \frac{db}{d\mu}$ denota o produto entre $\tilde{\phi}_\mu(a)$ e $\tilde{\phi}_\mu(b)$.

Note que $\langle \cdot, \cdot \rangle_\mu$ é um produto interno em $\mathcal{S}(\mathcal{X})$. A partir dele, podemos definir uma métrica riemanniana em $\mathcal{M}_+(\mathcal{X})$ que é, de fato, a métrica de Fisher.

Definição 9 (Métrica de Fisher). (AY *et al.*, 2017) Seja $\mu \in \mathcal{M}_+(\mathcal{X})$. Dados $A = (\mu, a)$, $B = (\mu, b) \in T_\mu \mathcal{M}_+(\mathcal{X})$, considere

$$g_\mu(A, B) := g(\mu)(A, B) := \langle a, b \rangle_\mu. \quad (3.5)$$

A métrica riemanniana g tal que $g|_{T_\mu \mathcal{M}_+(\mathcal{X})} = g(\mu)$ é chamada métrica de Fisher em $\mathcal{M}_+(\mathcal{X})$.

Note que, nesse contexto, 3.5 coincide com 2.6, mesmo os elementos de $\mathcal{M}_+(\mathcal{X})$ não sendo, necessariamente, normalizados sobre⁷ \mathcal{X} . Dito isso, podemos induzir a métrica de Fisher em $\mathcal{P}_+(\mathcal{X})$ tratando $\mathcal{P}_+(\mathcal{X})$ como subvariedade riemanniana de $\mathcal{M}_+(\mathcal{X})$, ou seja, a métrica de Fisher em $\mathcal{P}_+(\mathcal{X})$ pode ser definida como $\iota^* g$, que também denotaremos por g . Desse modo, em $\mathcal{P}_+(\mathcal{X})$, a matriz de informação de Fisher tem entradas dadas por⁸

$$\begin{aligned} G_{ij}(\xi) &= g_\xi(\partial_i(\xi), \partial_j(\xi)) \\ &= \langle \delta^i - \delta^{n+1}, \delta^j - \delta^{n+1} \rangle_\xi \\ &= \langle \delta^i, \delta^j \rangle_\xi + \langle \delta^{n+1}, \delta^{n+1} \rangle_\xi - \langle \delta^i, \delta^{n+1} \rangle_\xi - \langle \delta^{n+1}, \delta^j \rangle_\xi, \end{aligned}$$

⁷ Por normalizados, entende-se que $\sum_{i \in I} \mu_i = 1$ para todo $\mu \in \mathcal{M}_+(\mathcal{X})$.

⁸ Fica implícito que, quando escrevemos $\partial_i(\xi)$, estamos nos referindo ao vetor $(\xi, \partial_i(\xi)) \in T_\xi \mathcal{P}_+(\mathcal{X})$ para $i \in J$.

ou seja,

$$G_{ij}(\xi) = \begin{cases} \frac{1}{\xi_i} + \frac{1}{\xi_{n+1}} & , \text{ se } i = j \\ \frac{1}{\xi_{n+1}} & , \text{ caso contrário,} \end{cases}$$

cuja inversa consiste na matriz de entradas

$$G^{ij}(\xi) = \begin{cases} \xi_i(1 - \xi_i) & , \text{ se } i = j \\ -\xi_i\xi_j & , \text{ caso contrário.} \end{cases}$$

Ademais, vamos supor que queiramos estudar variedades mergulhadas em $\mathcal{M}_+(\mathcal{X})$ sob a óptica da métrica de Fisher, transferindo para elas, grosso modo, o caráter estatístico de $\mathcal{M}_+(\mathcal{X})$. Nesse contexto, podemos obter a métrica induzida nela pelo mergulho. Sejam, então, M uma variedade diferenciável de dimensão n e $m : M \rightarrow \mathcal{M}_+(\mathcal{X})$ um mergulho dado por $m(\zeta) = \sum_{i \in I} p_i(\zeta) \delta^i$. Fazendo o *pullback* da métrica de Fisher através desse mergulho, obtemos, para todo $A, B \in T_\zeta M$,

$$\begin{aligned} \tilde{g}(\zeta)(A, B) &:= m^* g(\zeta)(A, B) \\ &= g_{m(\zeta)} \left(\frac{\partial m}{\partial A}, \frac{\partial m}{\partial B} \right) \\ &= \sum_{i \in I} m_i(\zeta) \frac{\partial \ln m_i}{\partial A}(\zeta) \frac{\partial \ln m_i}{\partial B}(\zeta). \end{aligned}$$

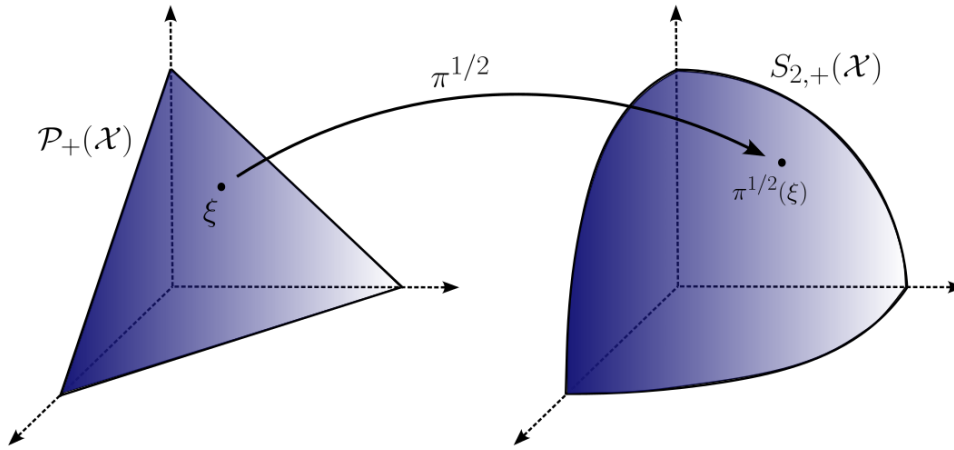
Note que a expressão anterior apresenta um formato próximo ao de 2.6, de maneira que a métrica de Fisher pode ser interpretada como uma métrica riemanniana em M . O que se encontra por trás disso são as diferenças entre uma abordagem intrínseca e uma extrínseca para se estudar variedades sob a perspectiva da relação entre geometria e estatística, que é fundante da geometria da informação. Por um lado, no que foi apresentado acima, a métrica de Fisher foi induzida em M pelo mergulho, enquanto que, por outro, ela poderia ter sido definida desde o início em M , forma que será mostrada com mais detalhes em 4.

Mostremos, agora, que $(\mathcal{P}_+(\mathcal{X}), g)$ possui uma estrutura familiar do ponto de vista geométrico. Para tal, considere o setor da esfera de raio 2 em $\mathcal{F}(\mathcal{X})$, dado por

$$S_{2,+}(\mathcal{X}) := \left\{ f \in \mathcal{F}(\mathcal{X}) \mid f(x_i) > 0, i \in I, \sum_{i \in I} f(x_i)^2 = 4 \right\} \subset \mathcal{F}(\mathcal{X}).$$

Considere, também, o difeomorfismo

$$\pi^{1/2} : \mathcal{P}_+(\mathcal{X}) \rightarrow S_{2,+}(\mathcal{X}), \quad \xi = \sum_{i \in I} \xi_i \delta^i \mapsto 2 \sum_{i \in I} \sqrt{\xi_i} e_i.$$

Figura 5 – Representação gráfica do difeomorfismo $\pi^{1/2}$ entre $\mathcal{P}_+(\mathcal{X})$ e $S_{2,+}(\mathcal{X})$ 

Fonte: Adaptada de Ay *et al.* (2017, p 33).

Vamos mostrar que $\pi^{1/2}$ é uma isometria entre $(\mathcal{P}_+(\mathcal{X}), g)$ e $(S_{2,+}(\mathcal{X}), \hat{g})$, onde \hat{g} é a métrica usual⁹ $\langle \cdot, \cdot \rangle$ de $\mathcal{F}(\mathcal{X}) \cong \mathbb{R}^{n+1}$ induzida em $S_{2,+}(\mathcal{X})$.

Proposição 1. (AY *et al.*, 2017) O mapa $\pi^{1/2}$ é uma isometria entre $\mathcal{P}_+(\mathcal{X})$ com a métrica de Fisher g e $S_{2,+}(\mathcal{X})$ com o produto escalar canônico induzido de $\mathcal{F}(\mathcal{X})$, ou seja,

$$\left\langle \frac{\partial \pi^{1/2}}{\partial A}(\xi), \frac{\partial \pi^{1/2}}{\partial B}(\xi) \right\rangle = g_\xi(A, B), \quad A, B \in T_\xi \mathcal{P}_+(\mathcal{X}).$$

Em outras palavras, a métrica de Fisher coincide com o *pullback* da métrica padrão em $\mathcal{F}(\mathcal{X})$ pelo mapa $\pi^{1/2}$.

Demonstração. Sejam $A = (\xi, a)$, $B = (\xi, b) \in T_\xi \mathcal{P}_+(\mathcal{X})$. Assim, temos que

$$\begin{aligned} \left\langle \frac{\partial \pi^{1/2}}{\partial A}(\xi), \frac{\partial \pi^{1/2}}{\partial B}(\xi) \right\rangle &= \left\langle \frac{d}{dt} \Big|_{t=0} \pi^{1/2}(\xi + ta), \frac{d}{dt} \Big|_{t=0} \pi^{1/2}(\xi + tb) \right\rangle \\ &= \sum_{i \in I} \frac{a_i}{\sqrt{\xi_i}} \frac{b_i}{\sqrt{\xi_i}} \\ &= \sum_{i \in I} \frac{a_i b_i}{\xi_i} \xi_i \\ &= g_\xi(A, B). \end{aligned}$$

□

Podemos, desse modo, unir essa proposição ao conhecido fato da geometria riemanniana de que a esfera $S_2 \subset \mathcal{F}(\mathcal{X})$ de raio 2 munida da métrica \hat{g} possui curvatura constante igual a $1/4$. Consequentemente, observamos que $\mathcal{P}_+(\mathcal{X})$ apresenta essa mesma curvatura.

⁹ Ou seja, para todo $f = \sum_{i \in I} f^i e_i$ e $g = \sum_{j \in I} g^j e_j$ elementos de $\mathcal{F}(\mathcal{X})$, a métrica $\langle \cdot, \cdot \rangle$ considerada em $\mathcal{F}(\mathcal{X})$ é $\langle f, g \rangle = \sum_{k \in I} f^k g^k$.

Ademais, como foi citado em 2.2.2, entre as opções para se considerar a relevância da métrica de Fisher, encontra-se o Teorema de Caracterização de Chentsov. Ele nos garante que essa é a única métrica (a menos de redimensionamento) que é invariante sob determinados mapas importantes, chamados de morfismos de Markov, sendo, então, a métrica destacada no contexto de modelos estatístico (DOWTY, 2018). Na definição de tais mapas, aparece o conjunto das medidas não negativas e normalizadas sobre \mathcal{X} , dado por

$$\mathcal{P}(\mathcal{X}) := \left\{ \xi \in \mathcal{S}(\mathcal{X}) \mid \xi_i \geq 0, \sum_{i \in I} \xi_i = 1 \right\}.$$

Dito isso, vamos os definir.

Definição 10 (Núcleo de Markov e morfismo de Markov). (AY *et al.*, 2017) Sejam $\mathcal{X} = \{x_1, x_2, \dots, x_{n+1}\}$ e $\mathcal{X}' = \{x'_1, x'_2, \dots, x'_{m+1}\}$ dois conjuntos finitos não vazios, com $n \leq m$. Definimos um núcleo de Markov K como um mapa $K : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X}')$ dado por

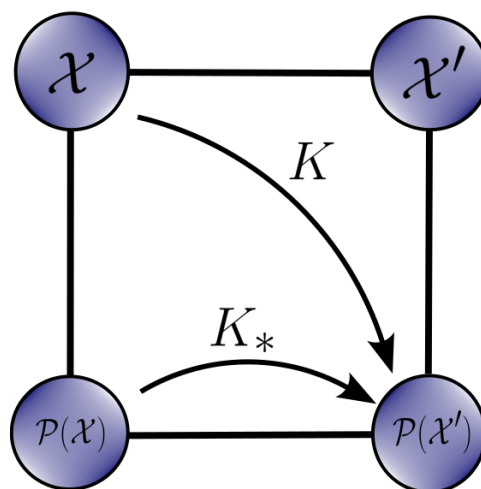
$$K(x_i) = K^i := \sum_{i'=1}^{m+1} K_{i'}^i \delta^{i'}.$$

Cada núcleo de Markov K induz um segundo mapa, chamado de morfismo de Markov induzido por K , denotado por K_* e definido como

$$K_* : \mathcal{P}(\mathcal{X}) \rightarrow \mathcal{P}(\mathcal{X}'), \quad \xi = \sum_{i \in I} \xi_i \delta^i \mapsto \sum_{i \in I} \xi_i K^i.$$

Dizemos que K é congruente se existe uma partição A_i , com $i \in I$, de \mathcal{X}' tal que $K_{i'}^i > 0$ se, e somente se, $x'_i \in A_i$.

Figura 6 – Representação gráfica das relações entre \mathcal{X} , \mathcal{X}' , $\mathcal{P}(\mathcal{X})$ e $\mathcal{P}(\mathcal{X}')$ através de núcleos de Markov e morfismos de Markov



Fonte: Elaborada pelo autor.

Perceba que toda função¹⁰ $f : \mathcal{X} \rightarrow \mathcal{X}'$ induz um núcleo de Markov K^f dado por¹¹

¹⁰ Por função estamos nos referindo a funções determinísticas.

¹¹ Quando não houver ambiguidade, escreveremos $\delta^{f(i)}$ para representar $\delta^{\mathcal{J} \circ f(x_i)}$.

$K^f(x_i) = \delta^{\mathcal{J} \circ f(x_i)}$, para todo $i \in I$, onde $\mathcal{J} : \mathcal{X} \rightarrow I$ é a função índice dada por $\mathcal{J}(x_i) = i$ para todo $x_i \in \mathcal{X}$. Além disso, os núcleos congruente preservam distribuições de probabilidade positivas, de maneira que, se K é núcleo de Markov congruente e $\xi \in \mathcal{P}_+(\mathcal{X})$, então $K_*(\xi) \in \mathcal{P}_+(\mathcal{X}')$ (AY *et al.*, 2017).

Podemos notar que K_* age sobre os elementos de $\mathcal{P}_+(\mathcal{X})$ tal como se ele fosse um mapa linear entre¹² $\mathcal{S}(\mathcal{X})$ e $\mathcal{S}(\mathcal{X}')$. De fato, sejam $\xi, \nu \in \mathcal{P}_+(\mathcal{X})$ e $a, b \in \mathbb{R}$. Então,

$$\begin{aligned} K_*(a\xi + b\nu) &= \sum_{i \in I} (a\xi + b\nu)_i K^i \\ &= a \sum_{i \in I} \xi_i K^i + b \sum_{i \in I} \nu_i K^i \\ &= aK_*(\xi) + bK_*(\nu). \end{aligned}$$

Desse modo, podemos esperar que a diferencial de K_* em $\xi \in \mathcal{P}_+(\mathcal{X})$ seja da forma

$$d_\xi K_* : T_\xi \mathcal{P}_+(\mathcal{X}) \rightarrow T_{K_*(\xi)} \mathcal{P}_+(\mathcal{X}'), \quad (\xi, \nu - \xi) \mapsto (K_*(\xi), K_*(\nu) - K_*(\xi))$$

e, efetivamente, é. Com efeito, sejam $\xi \in \mathcal{P}_+(\mathcal{X})$ e $A = (\xi, \nu - \xi) \in T_\xi \mathcal{P}_+(\mathcal{X})$. Note que, para $\varepsilon > 0$ suficientemente pequeno, $\xi + t(\nu - \xi) \in \mathcal{P}_+(\mathcal{X})$ para $t \in (-\varepsilon, \varepsilon)$. Desse modo,

$$\begin{aligned} d_\xi K_*(A) &= \frac{\partial K_*}{\partial A}(\xi) \\ &= \left. \frac{d}{dt} \right|_{t=0} K_*(\xi + t(\nu - \xi)) \\ &= \left. \frac{d}{dt} \right|_{t=0} \left(\sum_{i \in I} \xi_i K^i + t \sum_{i \in I} (\nu - \xi)_i K^i \right) \\ &= K_*(\nu) - K_*(\xi). \end{aligned}$$

Assim, a partir disso, podemos enunciar o Teorema da Caracterização de Chentsov¹³.

Teorema 2 (Teorema de Caracterização de Chentsov). (AY *et al.*, 2017) Atribuímos a cada conjunto finito e não vazio \mathcal{X} uma métrica $h^\mathcal{X}$ em $\mathcal{P}_+(\mathcal{X})$. Se, para cada núcleo de Markov congruente $K : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X}')$, temos invariância no sentido

$$h^\mathcal{X}(\xi)(A, B) = h^{\mathcal{X}'}(K_*(\xi))(d_\xi K_*(A), d_\xi K_*(B)), \quad (3.6)$$

então existe uma constante $\alpha > 0$ tal que $h^\mathcal{X} = \alpha g^\mathcal{X}$ para todos os \mathcal{X} , onde $g^\mathcal{X}$ é a métrica de Fisher em $\mathcal{P}_+(\mathcal{X})$.

Demonstração. Primeiramente, seja $\sigma : \mathcal{X} \rightarrow \mathcal{X}$ permutação em \mathcal{X} e defina

$$c_\mathcal{X} := \frac{1}{n} \sum_{i \in J} \delta^i, \quad n := |\mathcal{X}|.$$

¹² Aplicamos essa escolha vocabular menos precisa pois estamos considerando mapas lineares como definidos entre espaços vetoriais.

¹³ Mesmo apresentando uma demonstração razoavelmente longa, decidimos a manter no corpo principal do trabalho, dada a importância desse teorema.

Denotando $(K^\sigma)_*$ por σ_* , temos que

$$\begin{aligned}\sigma_*(c_{\mathcal{X}}) &= (K^\sigma)_*(c_{\mathcal{X}}) \\ &= \sum_{i \in J} (c_{\mathcal{X}})_i \delta^{\sigma(i)} \\ &= \frac{1}{n} \sum_{i \in J} \delta^{\sigma(i)} \\ &= c_{\mathcal{X}},\end{aligned}$$

ou seja, $c_{\mathcal{X}}$ é invariante à esquerda por σ . Seja, então, $E_i := (c_{\mathcal{X}}, \delta^i - c_{\mathcal{X}}) \in T_{c_{\mathcal{X}}} \mathcal{P}_+(\mathcal{X})$, com $i \in J$. Desse modo,

$$\begin{aligned}d_{c_{\mathcal{X}}} \sigma_*(E_i) &= (\sigma_*(c_{\mathcal{X}}), \sigma_*(\delta^i) - \sigma_*(c_{\mathcal{X}})) \\ &= (c_{\mathcal{X}}, \delta^{\sigma(i)} - c_{\mathcal{X}}) \\ &= E_{\sigma(i)},\end{aligned}$$

onde, para tornar a notação mais sucinta, $\sigma(i)$ corresponde à composição $\mathcal{J} \circ \sigma$ aplicada a x_i . Agora, para cada $i, j \in J$, com $i \neq j$, tome σ a transposição de x_i e x_j , ou seja,

$$\sigma(k) = \begin{cases} i, & \text{se } k = j \\ j, & \text{se } k = i \\ k, & \text{caso contrário} \end{cases}$$

Calculando, então, $h_{ii}^{\mathcal{X}}(c_{\mathcal{X}})$ sob a hipótese da invariância 3.6, obtemos

$$\begin{aligned}h_{ii}^{\mathcal{X}}(c_{\mathcal{X}}) &= h_{c_{\mathcal{X}}}^{\mathcal{X}}(E_i, E_i) \\ &= h_{\sigma_*(c_{\mathcal{X}})}^{\mathcal{X}}(d_{c_{\mathcal{X}}} \sigma_*(E_i), d_{c_{\mathcal{X}}} \sigma_*(E_i)) \\ &= h_{c_{\mathcal{X}}}^{\mathcal{X}}(E_{\sigma(i)}, E_{\sigma(i)}) \\ &= h_{c_{\mathcal{X}}}^{\mathcal{X}}(E_j, E_j) \\ &= h_{jj}^{\mathcal{X}}(c_{\mathcal{X}}) =: f_1(n).\end{aligned}$$

Note que $f_1(n)$ depende apenas de n , dado que $c_{\mathcal{X}}$ depende apenas dele, também. Além disso, escrevendo a expressão 3.6 para $h_{ij}^{\mathcal{X}}(c_{\mathcal{X}})$, com $i \neq j$, observamos que esses termos da métrica coincidem entre si. Assim, definindo $f_2(n) := h_{ij}^{\mathcal{X}}(c_{\mathcal{X}})$, podemos perceber que

$$\begin{aligned}f_1(n) + (n-1)f_2(n) &= \sum_{j \in J} h_{ij}^{\mathcal{X}}(c_{\mathcal{X}}) \\ &= \sum_{j \in J} h^{\mathcal{X}}(E_i, E_j) \\ &= h^{\mathcal{X}}\left(E_i, \underbrace{\sum_{j \in J} E_j}_0\right), \\ &= 0.\end{aligned}$$

Consequentemente, $f_1(n) = (1 - n)f_2(n)$. Agora, sejam $A = (\xi, a)$, $B = (\xi, b) \in T_\xi \mathcal{P}_+(\mathcal{X})$. Então,

$$\begin{aligned} h_{c_{\mathcal{X}}}^{\mathcal{X}}(A, B) &= \sum_{i \in J} a_i b_i h_{ii}^{\mathcal{X}}(c_{\mathcal{X}}) + \sum_{\substack{i, j \in J \\ i \neq j}} a_i b_j h_{ij}^{\mathcal{X}}(c_{\mathcal{X}}) \\ &= f_1(n) \sum_{i \in J} a_i b_i + f_2(n) \sum_{\substack{i, j \in J \\ i \neq j}} a_i b_j \\ &= f_2(n) \left(-n \sum_{i \in J} a_i b_i + \sum_{i, j \in J} a_i b_j \right) \\ &= -f_2(n) \sum_{i \in J} \frac{a_i b_i}{\frac{1}{n}} \\ &= -f_2(n) g_{c_{\mathcal{X}}}^{\mathcal{X}}(A, B). \end{aligned}$$

Mostremos, agora, que $f_2(n)$ é, de fato, constante. Para tal, seja $\tilde{\mathcal{X}} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\}$ conjunto finito não vazio e defina $\mathcal{X}' := \mathcal{X} \times \tilde{\mathcal{X}}$. Considere a partição de \mathcal{X}' cujos elementos são conjuntos da forma $A_i := \{(x_i, \tilde{x}_j) \in \mathcal{X}' \mid 1 \leq j \leq k\}$. Desse modo, definimos o núcleo de Markov

$$K : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X}'), \quad x_i \mapsto K^i = \frac{1}{k} \sum_{j=1}^k \delta^{(i,j)},$$

cujo morfismo induzido é dado por

$$K_*(c_{\mathcal{X}}) = \frac{1}{nk} \sum_{i \in J} \sum_{j=1}^k \delta^{(i,j)} = c_{\mathcal{X}'}$$

Consequentemente,

$$\begin{aligned} d_{c_{\mathcal{X}}} K_*(E_i) &= (K_*(c_{\mathcal{X}}), K_*(\delta^i) - K_*(c_{\mathcal{X}})) \\ &= (c_{\mathcal{X}'}, K_*(\delta^i) - c_{\mathcal{X}'}) \\ &= \left(c_{\mathcal{X}'}, \frac{1}{k} \sum_{j=1}^k \left(\delta^{(i,j)} - \frac{1}{nk} \sum_{l \in J} \sum_{m=1}^k \delta^{(l,m)} \right) \right) \\ &= \frac{1}{k} \sum_{j=1}^k E'_{(i,j)}, \end{aligned}$$

onde $E'_{(i,j)} := (c_{\mathcal{X}'}, \delta^{(i,j)} - c_{\mathcal{X}'})$. Sejam, então, $r, s \in J$, com $r \neq s$. Assim, aplicando a invariância 3.6, podemos escrever

$$\begin{aligned} f_2(n) &= h_{c_{\mathcal{X}}}^{\mathcal{X}}(E_r, E_s) \\ &= h_{c_{\mathcal{X}'}}^{\mathcal{X}'} \left(\frac{1}{k} \sum_{l=1}^k E'_{(r,l)}, \frac{1}{k} \sum_{m=1}^k E'_{(s,m)} \right) \\ &= \frac{1}{k^2} \sum_{l,m=1}^k h_{c_{\mathcal{X}'}}^{\mathcal{X}'}(E'_{(r,l)}, E'_{(s,m)}) \\ &= f_2(nk). \end{aligned}$$

Invertendo a função de n e k , obtemos $f_2(k) = f_2(nk) = f_2(n)$, de maneira que fica mostrado que f_2 é independente da cardinalidade de \mathcal{X} . Além disso, como $h_{ii}^{\mathcal{X}}(c_{\mathcal{X}}) = -f_2(n)g_{ii}^{\mathcal{X}}(c_{\mathcal{X}})$ para todo $i \in J$, temos que $-f_2(n) > 0$. Denotando $-f_2(n)$ por α , vamos estender a igualdade $h^{\mathcal{X}} = \alpha g^{\mathcal{X}}$ para todos os pontos de $\mathcal{P}_+(\mathcal{X})$. Visando isso, considere $\xi \in \mathcal{P}_+(\mathcal{X})$ cujas coordenadas na base canônica de $\mathcal{S}(\mathcal{X})$ são racionais, ou seja,

$$\xi = \sum_{i \in J} \frac{k_i}{n} \delta^i, \quad \sum_{i \in J} k_i = n.$$

Sejam, também, $\widetilde{\mathcal{X}}_i$ conjuntos de cardinalidade k_i , com $i \in J$, e que denotaremos por $\widetilde{\mathcal{X}}_i := \{\widetilde{x}_1^i, \widetilde{x}_2^i, \dots, \widetilde{x}_{k_i}^i\}$. Assim, defina o conjunto

$$\mathcal{X}' := \bigcup_{i=1}^n \left(\{x_i\} \times \widetilde{\mathcal{X}}_i \right)$$

e considere o núcleo da Markov

$$K : \mathcal{X} \rightarrow \mathcal{P}_+(\mathcal{X}'), \quad x_i \mapsto \frac{1}{k_i} \sum_{j=1}^{k_i} \delta^{(i,j)}. \quad (3.7)$$

Note que

$$\begin{aligned} K_*(\xi) &= \sum_{i \in J} \frac{k_i}{n} \frac{1}{k_i} \sum_{j=1}^{k_i} \delta^{(i,j)} \\ &= \frac{1}{n} \sum_{i \in J} \sum_{j=1}^{k_i} \delta^{(i,j)} \\ &= c_{\mathcal{X}'} \end{aligned}$$

e que, dado $A = (\xi, a) \in \mathcal{P}_+(\mathcal{X})$,

$$d_{\xi} K_*(A) = (c_{\mathcal{X}'}, K_*(a)) = \left(c_{\mathcal{X}'}, \sum_{i \in J} \sum_{j=1}^{k_i} \frac{a_i}{k_i} \delta^{(i,j)} \right).$$

Desse modo, aplicando a invariância 3.6, obtemos, para todo $A = (\xi, a), B = (\xi, b) \in \mathcal{P}_+(\mathcal{X})$, que

$$\begin{aligned} h_{\xi}^{\mathcal{X}}(A, B) &= h_{K_*(\xi)}^{\mathcal{X}'}(d_{\xi} K_*(A), d_{\xi} K_*(B)) \\ &= \alpha g_{c_{\mathcal{X}'}}^{\mathcal{X}'}(d_{\xi} K_*(A), d_{\xi} K_*(B)) \\ &= \alpha \sum_{i \in J} n \frac{a_i b_i}{k_i^2} k_i \\ &= \alpha g_{\xi}^{\mathcal{X}}(A, B). \end{aligned}$$

Como 3.7 sempre pode ser construído para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ com coordenadas racionais, temos que a sequência de igualdades acima é verdadeira para todo ξ nessas condições. Como $h^{\mathcal{X}}$ e $\alpha g^{\mathcal{X}}$ coincidem nesses pontos e como \mathbb{Q} é denso em \mathbb{R} , temos que $h^{\mathcal{X}} = \alpha g^{\mathcal{X}}$. \square

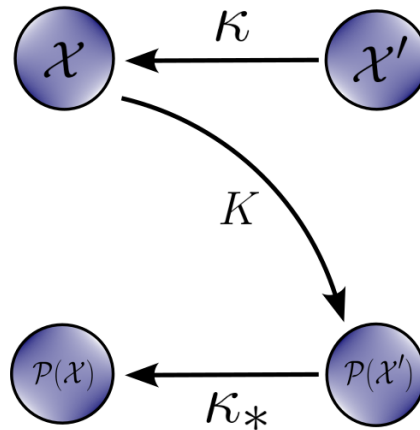
O teorema anterior nos mostra que a métrica de Fisher em modelos estatísticos sob conjuntos finitos é uma métrica destacada, uma vez que ela é a única, a menos de redimensi-onamentos, que é invariante por morfismos de Markov. Como esses morfismos são induzidos por núcleos de Markov, um questionamento razoável diz respeito ao significado de tais funções dentro da teoria estatística. Eles são, de maneira não muito precisa, consequência de construções advindas de estatísticas.

Definição 11 (Estatística). (SCHWACHHÖFER *et al.*, 2018) Dados dois conjuntos finitos não vazios $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ e $\mathcal{X}' = \{x'_1, x'_2, \dots, x'_m\}$, uma estatística entre eles é uma função $\kappa : \mathcal{X}' \rightarrow \mathcal{X}$. Já um núcleo de Markov $K : \mathcal{X}' \rightarrow \mathcal{P}(\mathcal{X})$ é dito congruente com respeito a κ se

$$\kappa_* K(x'_i) := \kappa_* \circ K(x'_i) = \delta^i,$$

para todo $i \in J$.

Figura 7 – Representação gráfica das relações entre \mathcal{X} , \mathcal{X}' , $\mathcal{P}(\mathcal{X})$ e $\mathcal{P}(\mathcal{X}')$ através de núcleos de Markov congruentes com respeito à estatística $\kappa : \mathcal{X}' \rightarrow \mathcal{X}$



Fonte: Elaborada pelo autor.

Assim, seja $K : \mathcal{X}' \rightarrow \mathcal{P}(\mathcal{X})$ um núcleo de Markov entre os conjuntos finitos não vazios $\mathcal{X} = \{x_1, x_2, \dots, x_n\}$ e $\mathcal{X}' = \{x'_1, x'_2, \dots, x'_m\}$ e suponha que exista $\kappa : \mathcal{X}' \rightarrow \mathcal{X}$ estatística tal que K é congruente em relação a ela. Tome, então, a partição de \mathcal{X}' cujos elementos são $A_i := \kappa^{-1}(x_i)$, chamada de partição de \mathcal{X}' induzida por κ . Como

$$\delta^{\kappa(j)} = K^{\kappa}(x'_j) = \sum_{l=1}^m (K^{\kappa})_l^j \delta^l,$$

temos, então, que

$$\begin{aligned} \kappa_* K(x_j) &= \sum_{i=1}^m K_i^j \sum_{l \in J} (K^{\kappa})_l^i \delta^l \\ &= \sum_{i=1}^m K_i^j \delta^{\kappa(i)} \\ &= \delta^j, \end{aligned}$$

onde a última igualdade advém da congruência em relação a κ . Queremos mostrar que K é congruente, então, por um lado, $m \geq n$ por hipótese. Por outro lado, dado $\tilde{\xi} \in \mathcal{P}(\mathcal{X}')$, temos que

$$\begin{aligned} \kappa_* \left(\tilde{\xi} \right) &= \sum_{j=1}^m \xi_j (K^\kappa)^j \\ &= \sum_{j=1}^m \xi_j \sum_{i \in J} (K^\kappa)_i^j \delta^i \\ &= \sum_{j=1}^m \xi_j \delta^{\kappa(j)}, \end{aligned}$$

ou seja, $n \geq m$ e, conseqüentemente, $n = m$. Posto isso, suponha que, para $x'_r \in \mathcal{X}'$, $\kappa(x'_r) = x_j$. Então, dada a igualdade das cardinalidades, a expressão $\kappa_* K(x_j) = \delta^j$ conduz a $K_r^i = 1 > 0$. Já se, para $x'_r \in \mathcal{P}(\mathcal{X}')$, tivermos que $K_r^j > 0$, então

$$K_r^j \delta^{\kappa(r)} + \sum_{\substack{i=1 \\ i \neq r}}^m K_i^j \delta^{\kappa(i)} = \delta^j,$$

onde fica claro que $\kappa(x'_r) = x_j$. Logo, vemos que K é congruente no sentido definido em 10. Em literaturas como (SCHWACHHÖFER *et al.*, 2018), dado um núcleo de Markov $K : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X}')$, a existência de uma estatística tal que ele é congruente a ela é tomada como a definição de congruência para K . O que mostramos acima é que essa definição induz aquela que havíamos apresentado, porém, exige a suposição de que $|\mathcal{X}'| \geq |\mathcal{X}|$, que compõe parte de 10. Agora, dada $\kappa : \mathcal{X}' \rightarrow \mathcal{X}$, com $|\mathcal{X}'| \geq |\mathcal{X}|$, considere a partição de \mathcal{X}' induzida por κ e $K : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X}')$ um núcleo de Markov congruente com respeito a essa partição¹⁴. Sabemos que

$$\kappa_* K(x_j) = \sum_{i=1}^m K_i^j \delta^{\kappa(i)}, \quad \sum_{i=1}^m K_i^j = 1,$$

o que implica que, para $x'_{r_1}, x'_{r_2}, \dots, x'_{r_k} \in \mathcal{X}'$, $K_{r_l}^j > 0$, com $l = 1, 2, \dots, k$. Da congruência de K , obtemos, então, que

$$\kappa_* K(x_j) = \sum_{i=1}^k K_{r_i}^j \delta^j = \delta^j.$$

Desse modo, observamos que, por outro lado, a definição 10 para congruência induz diretamente a definição via estatística. Isso abre precedentes para que a relação entre essas definições possa ser analisada através de estatísticas destacadas, chamadas de estatísticas suficientes. Vamos as definir para conjuntos não vazios finitos, que consistem no caso que envolve a discussão acima, porém a definição abaixo continua válida para subconjuntos de \mathbb{R}^n e para conjuntos discreto de cardinalidade infinita contável.

¹⁴ Note que esse núcleo existe. Por exemplo, tome $K : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X}')$ dado por

$$K(x_i) = \sum_{k \in A_i} \frac{1}{|A_i|} \delta^k$$

para todo $x_i \in \mathcal{X}$ e onde $k \in A_i$ denota os k tais que $x'_k \in A_i$.

Definição 12 (Estatística suficiente). (AMARI; NAGAOKA, 2000) Sejam \mathcal{X} , \mathcal{X}' conjuntos não vazios finitos e seja $\kappa : \mathcal{X} \rightarrow \mathcal{X}'$ uma estatística que transforma o valor da variável aleatória X em $Y = \kappa(X)$. Então, dada a distribuição $p(x; \xi)$ de X , isso determina a distribuição $q(y; \xi)$ que governa Y . Considerando

$$r(x; \xi) = \frac{p(x; \xi)}{q(\kappa(x); \xi)},$$

se $r(x; \xi)$ não depende de ξ para todo x , dizemos que κ é uma estatística suficiente.

Desse modo, para uma estatística suficiente $\kappa : \mathcal{X} \rightarrow \mathcal{X}'$, podemos escrever

$$p(x; \xi) = q(\kappa(x); \xi)r(x),$$

de maneira que fica evidente que $q(\kappa(x); \xi)$ compreende a dependência de $p(x; \xi)$ em ξ (AMARI; NAGAOKA, 2000). Em outras palavras, a suficiência de κ pode ser entendida como a característica dessa estatística que expressa o fato de que nenhuma outra estatística sobre a mesma amostra é capaz de fornecer mais informação acerca do parâmetro ξ (FISHER, 1922). Por meio disso, o Teorema de Chentsov pode ser pensado em termos de estatísticas, em especial, sob a perspectiva das estatísticas suficientes.

3.3 Conexões mistura e exponencial

Essa seção é dedicada a apresentar conexões lineares através das quais é possível escrever o análogo¹⁵ de *rank* 3 da métrica de Fisher. Começamos, então, definindo duas funções entre espaços tangentes de $\mathcal{M}_+(\mathcal{X})$.

Definição 13. (AY *et al.*, 2017) Dados $\mu, \tilde{\mu} \in \mathcal{M}_+(\mathcal{X})$, definimos os mapas

$$\begin{aligned} \tilde{\Pi}_{\mu, \tilde{\mu}}^{(m)} : T_{\mu} \mathcal{M}_+(\mathcal{X}) &\longrightarrow T_{\tilde{\mu}} \mathcal{M}_+(\mathcal{X}), & \tilde{\Pi}_{\mu, \tilde{\mu}}^{(e)} : T_{\mu} \mathcal{M}_+(\mathcal{X}) &\longrightarrow T_{\tilde{\mu}} \mathcal{M}_+(\mathcal{X}) \\ (\mu, a) &\longmapsto (\tilde{\mu}, a) & (\mu, a) &\longmapsto \left(\tilde{\mu}, \left(\tilde{\Phi}_{\tilde{\mu}}^{-1} \circ \tilde{\Phi}_{\mu} \right) (a) \right), \end{aligned}$$

onde $\tilde{\Phi}$ é dada por 3.4.

Dados dois campos $A, B \in \mathfrak{X}(\mathcal{M}_+(\mathcal{X}))$, seja $\gamma : (-\varepsilon, \varepsilon) \rightarrow \mathcal{M}_+(\mathcal{X})$ uma curva suave tal que $\gamma(0) = \mu$ e $\dot{\gamma}(0) = a_{\mu}$, com $A(\mu) = (\mu, a_{\mu})$. Representando $B(\gamma(t)) = (\gamma(t), b_{\gamma(t)})$, definimos a operação

$$\frac{\partial b}{\partial a_{\mu}}(\mu) := \lim_{t \rightarrow 0} \frac{1}{t} (b_{\gamma(t)} - b_{\mu}).$$

Observe que a expressão acima consiste em $A(\mu)$ agindo como derivação em cada função coordenada de B escrito em um referencial. Assim, vemos que ela é independente da escolha de γ . Posto isso, definamos a conexão mistura.

¹⁵ Estamos nos referindo ao tensor de Amari-Chentsov, que será apresentado em 3.4.

Definição 14 (Conexão mistura). (AY *et al.*, 2017) Chamamos de conexão mistura em $\mathcal{M}_+(\mathcal{X})$ ou m -conexão em $\mathcal{M}_+(\mathcal{X})$ o mapa

$$\begin{aligned}\tilde{\nabla}^{(m)} : \mathfrak{X}(\mathcal{M}_+(\mathcal{X})) \times \mathfrak{X}(\mathcal{M}_+(\mathcal{X})) &\longrightarrow \mathfrak{X}(\mathcal{M}_+(\mathcal{X})) \\ (A, B) &\longmapsto \tilde{\nabla}_A^{(m)} B,\end{aligned}$$

dado por

$$\tilde{\nabla}_A^{(m)} B(\mu) = \left(\mu, \frac{\partial b}{\partial a_\mu}(\mu) \right).$$

É fácil ver que $\tilde{\nabla}^{(m)}$ é, de fato, uma conexão linear em $\mathcal{M}_+(\mathcal{X})$. Além disso, podemos observar que $\tilde{\Pi}_{\mu, \gamma(t)}^{(m)}$ é operador transporte paralelo ao longo da curva suave $\gamma: (-\varepsilon, \varepsilon) \rightarrow \mathcal{M}_+(\mathcal{X})$ tal que $\gamma(0) = \mu$. Apresentemos, agora, a conexão exponencial.

Definição 15 (Conexão exponencial). (AY *et al.*, 2017) Chamamos de conexão exponencial em $\mathcal{M}_+(\mathcal{X})$ ou e -conexão em $\mathcal{M}_+(\mathcal{X})$ o mapa

$$\begin{aligned}\tilde{\nabla}^{(e)} : \mathfrak{X}(\mathcal{M}_+(\mathcal{X})) \times \mathfrak{X}(\mathcal{M}_+(\mathcal{X})) &\longrightarrow \mathfrak{X}(\mathcal{M}_+(\mathcal{X})) \\ (A, B) &\longmapsto \tilde{\nabla}_A^{(e)} B,\end{aligned}$$

dado por

$$\tilde{\nabla}_A^{(e)} B(\mu) = \left(\mu, \frac{\partial b}{\partial a_\mu}(\mu) - \left(\frac{da_\mu}{d\mu} \frac{db_\mu}{d\mu} \right) \mu \right). \quad (3.8)$$

Perceba que, assim como $\tilde{\nabla}^{(m)}$, a conexão exponencial configura-se como uma conexão linear em $\mathcal{M}_+(\mathcal{X})$. Além disso, é fácil ver que $\tilde{\Pi}_{\mu, \gamma(t)}^{(e)}$ é seu operador de transporte paralelo ao longo de γ tal qual $\tilde{\Pi}_{\mu, \gamma(t)}^{(m)}$ é para $\tilde{\nabla}^{(m)}$. Vale pontuar que em diversos trabalhos de geometria da informação¹⁶, a introdução das conexões $\tilde{\nabla}^{(m)}$ e $\tilde{\nabla}^{(e)}$ é feita partindo desses operadores de transportes paralelos. Contudo, como essa é uma abordagem menos comum em geometria riemanniana, decidimos as definir sem fazer uso deles, porém demarcando quem eles são. Isso justifica termos definido esses operadores mesmo não os usando na definição de $\tilde{\nabla}^{(m)}$ e $\tilde{\nabla}^{(e)}$.

Dito isso e apresentadas $\tilde{\nabla}^{(m)}$ e $\tilde{\nabla}^{(e)}$, chamamos de m -geodésicas e e -geodésicas as geodésicas com respeito a, respectivamente, $\tilde{\nabla}^{(m)}$ e $\tilde{\nabla}^{(e)}$. Usaremos os superíndices (m) e (e) para nos referirmos às grandezas que dizem respeito a $\tilde{\nabla}^{(m)}$ e $\tilde{\nabla}^{(e)}$, nessa ordem. Calculemos então as m - e e -geodésicas.

Proposição 2. (AY *et al.*, 2017) As m - e e -geodésicas maximais com ponto inicial $\mu \in \mathcal{M}_+(\mathcal{X})$ e velocidade inicial $a \in T_\mu \mathcal{M}_+(\mathcal{X})$ são

$$\begin{aligned}\tilde{\gamma}^{(m)} : (t^-, t^+) &\longrightarrow \mathcal{M}_+(\mathcal{X}) \\ t &\longmapsto \mu + ta\end{aligned}$$

¹⁶ Como exemplos, podemos citar (NIELSEN, 2022), (AMARI, 2016) e o próprio (AY *et al.*, 2017).

com

$$t^- := -\min \left\{ \frac{\mu_i}{a_i} \mid i \in I, a_i > 0 \right\}, \quad t^+ := \min \left\{ \frac{\mu_i}{|a_i|} \mid i \in I, a_i < 0 \right\}$$

sob a convenção $\min \emptyset = +\infty$, e

$$\begin{aligned} \tilde{\gamma}^{(e)} : \mathbb{R} &\longrightarrow \mathcal{M}_+(\mathcal{X}) \\ t &\longmapsto \exp \left(t \frac{da}{d\mu} \right) \mu. \end{aligned}$$

Demonstração. Começemos pelo caso da m -conexão. Para encontrarmos as m -geodésicas, precisamos resolver a equação da geodésica. Assim, dado $\mu \in \mathcal{M}_+(\mathcal{X})$, por um lado, temos que

$$g_\mu \left(\tilde{\nabla}_i^{(m)} \partial_j(\mu), \partial_k \right) = g_\mu \left(\sum_{l \in I} \tilde{\Gamma}_{ij}^{(m)l}(\mu) \partial_l, \partial_k \right) = \frac{\tilde{\Gamma}_{ij}^{(m)k}(\mu)}{\mu_k},$$

onde¹⁷ $\tilde{\nabla}_i^{(m)} := \tilde{\nabla}_{\partial_i}^{(m)}$. Por outro lado,

$$g_\mu \left(\tilde{\nabla}_i^{(m)} \partial_j(\mu), \partial_k \right) = g_\mu \left(\sum_{l \in I} \frac{d}{dt} \Big|_{t=0} (\partial_j)_{\gamma(t), l} \delta^l, \partial_k \right) = 0.$$

Desse modo,

$$\tilde{\Gamma}_{ij}^{(m)k} \equiv 0$$

para todo $i, j, k \in I$ e, conseqüentemente, a equação da geodésica se resume a

$$\ddot{\gamma}(t) = 0,$$

cuja solução, impondo as condições $\gamma(0) = \mu \in \mathcal{M}_+(\mathcal{X})$ e $\dot{\gamma}(0) = a \in T_\mu \mathcal{M}_+(\mathcal{X})$, é

$$\tilde{\gamma}^{(m)}(t) = \mu + ta.$$

Como os elementos de $\mathcal{M}_+(\mathcal{X})$ têm coordenadas estritamente positivas em $\mathcal{B}_{\mathcal{S}(\mathcal{X})}$, $\tilde{\gamma}^{(m)}$ está definida no intervalo maximal (t^-, t^+) , com t^- e t^+ dados de acordo com o enunciado.

Agora, para o caso da e -conexão, sabemos que

$$g_\mu \left(\tilde{\nabla}_i^{(e)} \partial_j(\mu), \partial_k \right) = \frac{\tilde{\Gamma}_{ij}^{(e)k}(\mu)}{\mu_k}.$$

Em contrapartida,

$$\begin{aligned} g_\mu \left(\tilde{\nabla}_i^{(e)} \partial_j(\mu), \partial_k \right) &= g_\mu \left(\frac{\partial (\partial_j)}{\partial (\partial_i)_\mu}(\mu) - \left(\frac{d(\partial_i)_\mu}{d\mu} \cdot \frac{d(\partial_i)_\mu}{d\mu} \right) \mu, \partial_k \right) \\ &= g_\mu \left(- \sum_{l \in I} \frac{1}{\mu_l} (\partial_i)_{\mu l} (\partial_j)_{\mu l} \delta^l, \partial_k \right) \\ &= - \frac{1}{\mu_k^2} \delta_{ij} \delta_{ik}. \end{aligned}$$

¹⁷ Omitimos o ponto de aplicação dos campos coordenados, uma vez que ele está subentendido, ou seja, ∂_i denota $\partial_i(\mu)$.

Desse modo, obtemos que

$$\tilde{\Gamma}_{ij}^{(e)k}(\mu) = -\frac{1}{\mu_k} \delta_{ij} \delta_{ik}$$

para todo $i, j, k \in I$ e para todo $\mu \in \mathcal{M}_+(\mathcal{X})$. Assim, a equação da geodésica se escreve como

$$\ddot{\gamma} - \frac{\dot{\gamma}^2}{\gamma} = 0,$$

cuja solução, impondo as mesmas condições que imprimimos para o caso da m -conexão, é simplesmente

$$\tilde{\gamma}^{(e)}(t) = \exp\left(t \frac{da}{d\mu}\right) \mu = \sum_{i \in I} \mu_i e^{t \frac{a_i}{\mu_i}} \delta^i.$$

Como a exponencial é função estritamente positiva, $\mathcal{M}_+(\mathcal{X})$ é geodesicamente completo em relação a $\tilde{\nabla}^{(e)}$. \square

Corolário 1. (AY *et al.*, 2017) Os mapas exponenciais $\tilde{\text{exp}}^{(m)}$ e $\tilde{\text{exp}}^{(e)}$ respectivos a $\tilde{\nabla}^{(m)}$ e $\tilde{\nabla}^{(e)}$ são dados por

$$\begin{aligned} \tilde{\text{exp}}^{(m)} : T &\longrightarrow \mathcal{M}_+(\mathcal{X}) \\ (\mu, a) &\longmapsto \mu + a, \end{aligned}$$

com

$$T := \{(\mu, \tilde{\mu} - \mu) \in T\mathcal{M}_+(\mathcal{X}) \mid \mu, \tilde{\mu} \in \mathcal{M}_+(\mathcal{X})\},$$

e

$$\begin{aligned} \tilde{\text{exp}}^{(e)} : T\mathcal{M}_+(\mathcal{X}) &\longrightarrow \mathcal{M}_+(\mathcal{X}) \\ (\mu, a) &\longmapsto \exp\left(\frac{da}{d\mu}\right) \mu. \end{aligned}$$

Demonstração. Para o caso da m -conexão, dado $(\mu, a) \in T$, considere $\tilde{\gamma}^{(m)}$ com $\tilde{\gamma}^{(m)}(0) = \mu$ e $\dot{\tilde{\gamma}}^{(m)}(0) = a$ e tome $t = 1$, ou seja,

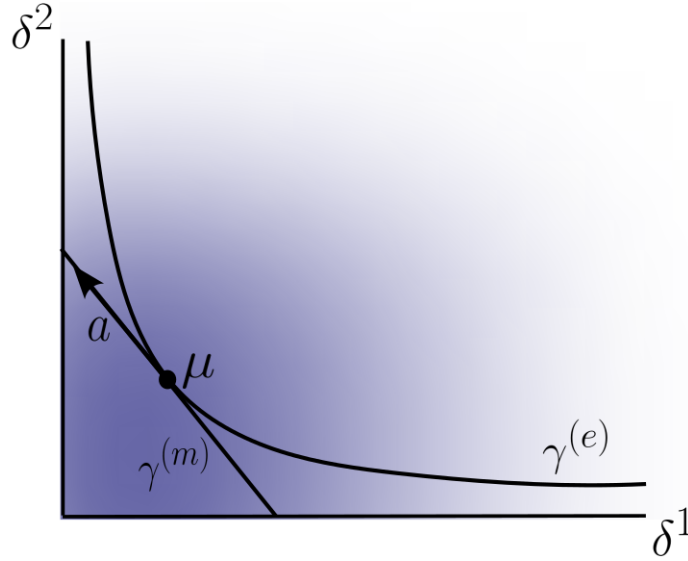
$$\tilde{\text{exp}}^{(m)}(\mu, a) = \tilde{\gamma}^{(m)}(1) = \mu + a.$$

Através dessa expressão, é possível notar que, se $(\mu, a) \notin T$ com $\mu \in \mathcal{M}_+(\mathcal{X})$, então $a + \mu \leq 0$ e, por consequência, $\tilde{\text{exp}}^{(m)}(\mu, a)$ não está definido. Assim, $\tilde{\text{exp}}^{(m)}$ está definido em T . Reproduzindo analogamente o que foi feito acima para o caso da e -conexão, obtemos que

$$\tilde{\text{exp}}^{(e)}(\mu, a) = \tilde{\gamma}^{(e)}(1) = \exp\left(\frac{da}{d\mu}\right) \mu,$$

que está definido para todo $(\mu, a) \in T\mathcal{M}_+(\mathcal{X})$. \square

Figura 8 – Exemplo de m - e e -geodésica em $\mathcal{M}_+(\mathcal{X})$ no caso de $\dim \mathcal{M}_+(\mathcal{X}) = 2$. Como $\tilde{\nabla}^{(m)}$ possui símbolos de Christoffel nulos, suas geodésicas são retas em $\mathcal{M}_+(\mathcal{X})$, enquanto as de $\tilde{\nabla}^{(e)}$ são exponenciais



Fonte: Elaborada pelo autor.

Podemos observar que $\tilde{\nabla}^{(m)}$ e $\tilde{\nabla}^{(e)}$ podem ser induzidas em $\mathcal{P}_+(\mathcal{X})$. Denotando a m -conexão em $\mathcal{P}_+(\mathcal{X})$ por $\nabla^{(m)}$, temos que, para todo $A, B \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$ e para todo $\xi \in \mathcal{P}_+(\mathcal{X})$,

$$\nabla_A^{(m)} B(\xi) = \tilde{\nabla}_A^{(m)} \tilde{B}(\xi),$$

onde $\tilde{A}, \tilde{B} \in \mathfrak{X}(U)$ são, respectivamente, extensões suaves de A e B a um aberto U de $\mathcal{M}_+(\mathcal{X})$ contendo $\mathcal{P}_+(\mathcal{X})$. Vamos mostrar que essas extensões realmente existem.

Proposição 3. Seja $A : \mathcal{P}_+(\mathcal{X}) \rightarrow M$ função suave entre variedades. Então, existem U aberto de $\mathcal{M}_+(\mathcal{X})$ contendo $\mathcal{P}_+(\mathcal{X})$ e $\tilde{A} : U \rightarrow M$ função suave tal que $\tilde{A}|_{\mathcal{P}_+(\mathcal{X})} = A$.

Demonstração. Como A é suave, para cada $\xi \in \mathcal{P}_+(\mathcal{X})$, existem U_ξ aberto de $\mathcal{M}_+(\mathcal{X})$ contendo ξ e $\tilde{A}_\xi : U_\xi \rightarrow M$ função suave tal que $\tilde{A}_\xi|_{U_\xi \cap \mathcal{P}_+(\mathcal{X})} = A$. Assim, $\{U_\xi \mid \xi \in \mathcal{P}_+(\mathcal{X})\}$ é uma cobertura aberta para $\mathcal{P}_+(\mathcal{X})$. Seja, então, $\{\varphi_\xi\}_{\xi \in \mathcal{P}_+(\mathcal{X})}$ partição da unidade subordinada a essa cobertura. Para cada $\mu \in \mathcal{M}_+(\mathcal{X})$, defina

$$\mathcal{C}_\mu := \{\xi \in \mathcal{P}_+(\mathcal{X}) \mid \mu \in \text{supp}(\varphi_\xi)\}.$$

Defina, também,

$$U := \bigcup_{\xi \in \mathcal{P}_+(\mathcal{X})} U_\xi$$

e a função

$$\tilde{A} : U \rightarrow M, \quad \mu \mapsto \sum_{\xi \in \mathcal{C}_\mu} \varphi_\xi(\mu) \tilde{A}_\xi(\mu).$$

É imediato que \tilde{A} é suave e, além disso, para todo $\tilde{\xi} \in \mathcal{P}_+(\mathcal{X})$,

$$\tilde{A}(\tilde{\xi}) = \sum_{\xi \in \mathcal{C}_{\tilde{\xi}}} \varphi_{\xi}(\tilde{\xi}) \tilde{A}_{\xi}(\tilde{\xi}) = \sum_{\xi \in \mathcal{C}_{\tilde{\xi}}} \varphi_{\xi}(\tilde{\xi}) A_{\xi}(\tilde{\xi}) = A(\tilde{\xi}),$$

ou seja, $\tilde{A}|_{\mathcal{P}_+(\mathcal{X})} = A$. □

Assim, mostrada a existência das extensões de A e B campos de vetores suaves de $\mathcal{P}_+(\mathcal{X})$, vamos demonstrar que $\nabla_A^{(m)} B \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$. Sendo \tilde{A}, \tilde{B} extensões suaves, respectivamente, de A e B a um aberto $U \supset \mathcal{P}_+(\mathcal{X})$ de $\mathcal{M}_+(\mathcal{X})$, temos que

$$\begin{aligned} \nabla_A^{(m)} B(\mu) &= \tilde{\nabla}_{\tilde{A}}^{(m)} \tilde{B}(\mu) \\ &= \left(\mu, \frac{\partial \tilde{b}(\mu)}{\partial \tilde{a}_{\mu}} \right) \\ &= \left(\mu, \sum_{i \in I} \frac{d}{dt} \Big|_{t=0} \left(\tilde{b}_{\gamma(t)} \right)_i \delta^i \right) \end{aligned}$$

Como

$$\begin{aligned} \sum_{i \in I} \frac{d}{dt} \Big|_{t=0} \left(\tilde{b}_{\gamma(t)} \right)_i &= \frac{d}{dt} \Big|_{t=0} \left[\sum_{i \in I} \left(\tilde{b}_{\gamma(t)} \right)_i \right] \\ &= \lim_{t \rightarrow 0} \frac{\sum_{i \in I} \left(\tilde{b}_{\gamma(t)} \right)_i - \sum_{i \in I} \left(\tilde{b}_{\gamma(0)} \right)_i}{t} \\ &= \lim_{t \rightarrow 0} \frac{\sum_{i \in I} \left(b_{\gamma(t)} \right)_i - \sum_{i \in I} \left(b_{\gamma(0)} \right)_i}{t} \\ &= \frac{d}{dt} \Big|_{t=0} \underbrace{\left[\sum_{i \in I} \left(b_{\gamma(t)} \right)_i \right]}_0 \\ &= 0, \end{aligned}$$

concluimos que $\nabla_A^{(m)} B \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$. É fácil notar que a suavidade de $\nabla_A^{(m)} B$ advém da suavidade das suas entradas. Além disso, o cálculo acima deixa claro que $\nabla_A^{(m)} B$ não depende das extensões tomadas para A e B , tampouco do aberto U . Desse modo, podemos definir a conexão mistura $\nabla^{(m)}$ em $\mathcal{P}_+(\mathcal{X})$, grosso modo, como a restrição de $\tilde{\nabla}^{(m)}$ a $T\mathcal{P}_+(\mathcal{X})$. Esse processo não pode ser realizado, de maneira geral, para $\tilde{\nabla}^{(e)}$ devido ao termo $\left(\frac{da_{\mu}}{d\mu} \frac{db_{\mu}}{d\mu} \right) \mu$ em 3.8. A fim de lidarmos com esse contratempo, para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $A, B \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$, projetamos $\tilde{\nabla}_A^{(e)} \tilde{B}(\xi)$ em $T_{\xi} \mathcal{P}_+(\mathcal{X})$, onde \tilde{A}, \tilde{B} são extensões suaves de A e B , respectivamente, definidas em um aberto $U \supset \mathcal{P}_+(\mathcal{X})$ de $\mathcal{M}_+(\mathcal{X})$. Para tal, vamos obter uma expressão para essa projeção.

Proposição 4. (AY *et al.*, 2017) Sendo $\mathcal{M}_+(\mathcal{X})$ munida com a métrica de Fisher, dado

$\mu \in \mathcal{M}_+(\mathcal{X})$, para todo $a \in \mathcal{S}(\mathcal{X})$, a é decomposto em $a = \Pi_\mu^\top a + \Pi_\mu^\perp a$, com

$$\begin{aligned}\Pi_\mu^\perp a &= \sum_{i \in I} \left(a_i - \mu_i \sum_{j \in I} a_j \right) \delta^i \in \mathcal{S}_0(\mathcal{X}), \\ \Pi_\mu^\top a &= \sum_{i \in I} \left(\mu_i \sum_{j \in I} a_j \right) \delta^i.\end{aligned}$$

Demonstração. Como a métrica de Fisher g é não degenerada, sabemos que existe $\mathcal{S}_0(\mathcal{X})^\perp$ tal que $\mathcal{S}_0(\mathcal{X}) \cap \mathcal{S}_0(\mathcal{X})^\perp = \{0\}$, de maneira que $a = \Pi_\mu^\top a + \Pi_\mu^\perp a$ para todo $a \in \mathcal{S}(\mathcal{X})$. Afirmamos que $\mathcal{S}_0(\mathcal{X})^\perp = \mathbb{R} \cdot \mu := \{\lambda \mu \mid \lambda \in \mathbb{R}\}$. De fato, seja $b \in \mathcal{S}_0(\mathcal{X})$. Então, para todo $\mu \in \mathcal{M}_+(\mathcal{X})$,

$$g_\mu((\mu, b), (\mu, c)) = 0 \Leftrightarrow \sum_{i \in I} \frac{1}{\mu_i} b_i c_i = 0 \Leftrightarrow c_i = \lambda \mu_i, \lambda \in \mathbb{R}.$$

Assim, para todo $\mu \in \mathcal{M}_+(\mathcal{X})$, $a \in \mathcal{S}(\mathcal{X})$ e $\lambda \in \mathbb{R}$,

$$\begin{aligned}\Pi_\mu^\perp a &= \frac{g_\mu(a, \lambda \mu)}{g_\mu(\lambda \mu, \lambda \mu)} (\lambda \mu) \\ &= g_\mu(a, \mu) \mu \\ &= \sum_{i \in I} \left(\mu_i \sum_{j \in I} a_j \right) \delta^i.\end{aligned}$$

Consequentemente,

$$\begin{aligned}\Pi_\mu^\top a &= a - \Pi_\mu^\perp a \\ &= \sum_{i \in I} a_i \delta^i - \sum_{i \in I} \left(\mu_i \sum_{j \in I} a_j \right) \delta^i \\ &= \sum_{i \in I} \left(a_i - \mu_i \sum_{j \in I} a_j \right) \delta^i.\end{aligned}$$

□

Isso posto, dados $A, B \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$ e $\xi \in \mathcal{P}_+(\mathcal{X})$, considere \tilde{A}, \tilde{B} extensões suaves de A e B , respectivamente, a um aberto $U \supset \mathcal{P}_+(\mathcal{X})$ de $\mathcal{M}_+(\mathcal{X})$. Assim, defina a e -conexão em $\mathcal{P}_+(\mathcal{X})$ por

$$\nabla_A^{(e)} B(\xi) = \Pi_\xi^\top \tilde{\nabla}_{\tilde{A}}^{(e)} \tilde{B}(\xi) \in T_\xi \mathcal{P}_+(\mathcal{X}).$$

Escrevendo a expressão acima em coordenadas, obtemos

$$\begin{aligned}\nabla_A^{(e)} B(\xi) &= \left(\xi, \sum_{i \in I} \left[\frac{\partial \tilde{b}_i}{\partial \tilde{a}_\xi}(\xi) - \frac{1}{\xi_i} \tilde{a}_{\xi_i} \tilde{b}_{\xi_i} - \xi_i \left(\sum_{j \in I} \frac{\partial \tilde{b}_j}{\partial \tilde{a}_\xi}(\xi) - \frac{1}{\xi_j} \tilde{a}_{\xi_j} \tilde{b}_{\xi_j} \right) \right] \delta^i \right) \\ &= \left(\xi, \sum_{i \in I} \left[\frac{\partial b_i}{\partial a_\xi}(\xi) - \frac{1}{\xi_i} a_{\xi_i} b_{\xi_i} + \xi_i \left(\sum_{j \in I} \frac{1}{\xi_j} a_{\xi_j} b_{\xi_j} \right) \right] \delta^i \right) \\ &= \left(\xi, \frac{\partial b}{\partial a_\xi}(\xi) - \left(\frac{da_\xi}{d\xi} \frac{db_\xi}{d\xi} \right) \xi + g_\xi(A_\xi, B_\xi) \xi \right).\end{aligned}$$

No cálculo acima, fica evidente que $\nabla_A^{(e)}B$ independe das extensões \tilde{A}, \tilde{B} tomadas. Por conseguinte, não faremos distinção lexográfica entre os campos A e B e suas extensões durante o cálculo de $\nabla_A^{(e)}B$. Agora, repetindo o que foi feito para as conexões mistura e exponencial em $\mathcal{M}_+(\mathcal{X})$, vamos calcular as m - e e -geodésicas.

Proposição 5. (AY *et al.*, 2017) As m - e e -geodésicas maximais com ponto inicial $\xi \in \mathcal{P}_+(\mathcal{X})$ e velocidade inicial $a \in T_\xi \mathcal{P}_+(\mathcal{X})$ são

$$\begin{aligned} \gamma^{(m)} : (t^-, t^+) &\longrightarrow \mathcal{P}_+(\mathcal{X}) \\ t &\longmapsto \xi + ta \end{aligned}$$

com

$$t^- := -\min \left\{ \frac{\eta_i}{a_i} \mid i \in I, a_i > 0 \right\}, \quad t^+ := \min \left\{ \frac{\eta_i}{|a_i|} \mid i \in I, a_i < 0 \right\}$$

e

$$\begin{aligned} \gamma^{(e)} : \mathbb{R} &\longrightarrow \mathcal{P}_+(\mathcal{X}) \\ t &\longmapsto \frac{\exp\left(t \frac{da}{d\xi}\right)}{\xi \left(\exp\left(t \frac{da}{d\xi}\right)\right)} \xi. \end{aligned}$$

Demonstração. Precisamos provar apenas o caso referente à e -conexão, uma vez que aquele relativo à m -conexão segue diretamente de 2. Assim, escrevendo a equação da geodésica para $\nabla^{(e)}$, temos

$$\nabla_{\dot{\gamma}(t)}^{(e)} \dot{\gamma}(t) = \left(\gamma(t), \sum_{i \in I} \left[\frac{\partial \dot{\gamma}_i}{\partial \dot{\gamma}(t)}(\gamma(t)) - \frac{\dot{\gamma}^2(\gamma(t))}{\gamma(t)} - \gamma(t) \left(\sum_{j \in I} \frac{\dot{\gamma}_j^2(\gamma(t))}{\gamma_j(t)} \right) \right] \delta^i \right) = (\gamma(t), 0)$$

ou¹⁸, equivalentemente,

$$\ddot{\gamma} - \frac{\dot{\gamma}^2}{\gamma} + \gamma \sum_{i \in I} \frac{\dot{\gamma}_i^2}{\gamma_i} = 0, \quad (3.9)$$

cuja solução, impondo as condições $\gamma(0) = \xi \in \mathcal{P}_+(\mathcal{X})$ e $\dot{\gamma}(0) = a \in T_\xi \mathcal{P}_+(\mathcal{X})$, é

$$\gamma^{(e)}(t) = \frac{\exp\left(t \frac{da}{d\xi}\right)}{\xi \left(\exp\left(t \frac{da}{d\xi}\right)\right)} \xi = \sum_{i \in I} \frac{\xi_i e^{t \frac{a_i}{\xi_i}}}{\sum_{j \in I} \xi_j e^{t \frac{a_j}{\xi_j}}} \delta^i. \quad (3.10)$$

De fato, primeiramente¹⁹,

$$\gamma(0) = \sum_{i \in I} \frac{\xi_i e^0}{\sum_{j \in I} \xi_j e^0} \delta^i = \sum_{i \in I} \xi_i \delta^i = \xi.$$

¹⁸ Para manter a notação mais sucinta, escrevemos $\nabla_{\dot{\gamma}(t)}^{(e)} \dot{\gamma}(t)$ para $\nabla_{\dot{\gamma}(\gamma(t))}^{(e)} \dot{\gamma}(\gamma(t))$, convenção que usaremos quando formos expressar a equação da geodésica em termos explícitos de alguma derivada covariante.

¹⁹ A fim de tornar a notação menos densa, nos cálculos que seguem até o fim desta demonstração, denotaremos $\gamma^{(e)}$ apenas por γ .

Calculando as coordenadas de $\dot{\gamma}(t)$ e $\ddot{\gamma}(t)$, temos que

$$\begin{aligned}\dot{\gamma}_i(t) &= \gamma_i(t) \left(\frac{a_i}{\xi_i} - \sum_{j \in I} \gamma_j(t) \frac{a_j}{\xi_j} \right), \\ \ddot{\gamma}_i(t) &= \gamma_i(t) \left(\frac{a_i}{\xi_i} - \sum_{j \in I} \gamma_j(t) \frac{a_j}{\xi_j} \right)^2 - \gamma_i(t) \sum_{j \in I} \dot{\gamma}_j(t) \frac{a_j}{\xi_j}.\end{aligned}\tag{3.11}$$

Desse modo, vemos que

$$\dot{\gamma}(0) = \gamma_i(0) \left(\frac{a_i}{\xi_i} - \sum_{j \in I} \gamma_j(0) \frac{a_j}{\xi_j} \right) = \xi_i \left(\frac{a_i}{\xi_i} - \sum_{i \in I} \xi_j \frac{a_j}{\xi_j} \right) = a_i.$$

Substituindo a expressão 3.11 no lado esquerdo de 3.9, obtemos

$$\begin{aligned}\left(\ddot{\gamma}(t) - \frac{\dot{\gamma}(t)}{\gamma(t)} + \gamma(t) \sum_{j \in I} \frac{\dot{\gamma}_j(t)}{\gamma_j(t)} \right)_i &= \ddot{\gamma}_i(t) - \frac{\dot{\gamma}_i^2(t)}{\gamma_i(t)} + \gamma_i(t) \sum_{j \in I} \frac{\dot{\gamma}_j(t)^2}{\gamma_j(t)} \\ &= \gamma_i(t) \left(\frac{a_i}{\xi_i} - \sum_{j \in I} \gamma_j(t) \frac{a_j}{\xi_j} \right)^2 - \gamma_i(t) \sum_{j \in I} \dot{\gamma}_j(t) \frac{a_j}{\xi_j} \\ &\quad - \gamma_i(t) \left(\frac{a_i}{\xi_i} - \sum_{j \in I} \gamma_j(t) \frac{a_j}{\xi_j} \right)^2 + \gamma_i(t) \sum_{j \in I} \dot{\gamma}_j(t) \frac{a_j}{\xi_j} \\ &= 0.\end{aligned}$$

Logo, as e -geodésicas em $\mathcal{P}_+(\mathcal{X})$ são da forma

$$\gamma^{(e)}(t) = \frac{\exp\left(t \frac{da}{d\xi}\right)}{\xi \left(\exp\left(t \frac{da}{d\xi}\right) \right)} \xi,$$

com $\xi \in \mathcal{P}_+(\mathcal{X})$ e $a \in \mathcal{S}_0(\mathcal{X})$. Por fim, observando 3.10, vemos que, para todo $t \in \mathbb{R}$, $\gamma^{(e)}$ é normalizada e tem todas as entradas positivas. \square

Corolário 2. (AY *et al.*, 2017) Os mapas exponenciais $\exp^{(m)}$ e $\exp^{(e)}$ respectivos a $\nabla^{(m)}$ e $\nabla^{(e)}$ são dados por

$$\begin{aligned}\exp^{(m)} : T &\longrightarrow \mathcal{P}_+(\mathcal{X}) \\ (\xi, a) &\longmapsto \xi + a,\end{aligned}$$

com

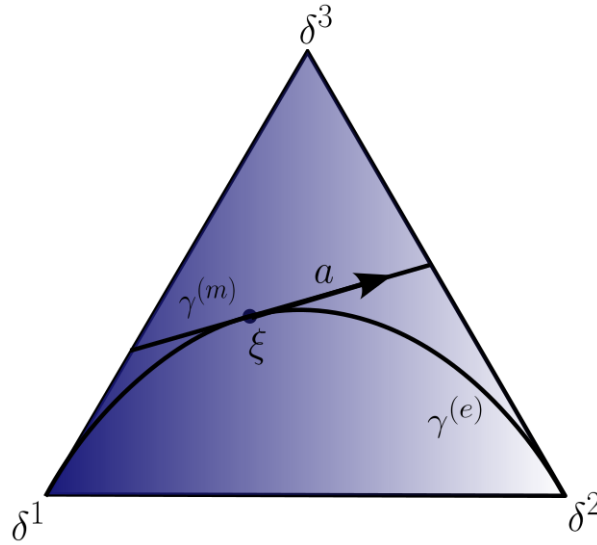
$$T := \left\{ \left(\xi, \tilde{\xi} - \xi \right) \in T \mathcal{P}_+(\mathcal{X}) \mid \xi, \tilde{\xi} \in \mathcal{P}_+(\mathcal{X}) \right\},$$

e

$$\begin{aligned}\exp^{(e)} : T \mathcal{P}_+(\mathcal{X}) &\longrightarrow \mathcal{P}_+(\mathcal{X}) \\ (\xi, a) &\longmapsto \frac{\exp\left(\frac{da}{d\xi}\right)}{\xi \left(\exp\left(\frac{da}{d\xi}\right) \right)} \xi.\end{aligned}$$

Demonstração. Novamente, é fácil ver que precisamos mostrar apenas o caso da e -conexão, em que basta tomar $t = 1$ em $\gamma^{(e)}(t)$ para obter a expressão para $\exp^{(e)}$ dada pelo enunciado e definida em todo $T\mathcal{P}_+(\mathcal{X})$. \square

Figura 9 – Exemplo de m - e e -geodésica em $\mathcal{P}_+(\mathcal{X})$ no caso de $\dim \mathcal{P}_+(\mathcal{X}) = 2$



Fonte: Adaptada de Ay *et al.* (2017, p 47).

3.4 O tensor de Amari-Chentsov

Apresentadas as conexões mistura e exponencial tanto em $\mathcal{M}_+(\mathcal{X})$ quanto $\mathcal{P}_+(\mathcal{X})$, podemos, a partir delas, expressar o tensor de Amari-Chentsov. Ele, junto à métrica de Fisher, constitui, grosso modo, uma estrutura por meio da qual autores como Steffen Lauritzen definem variedades estatísticas (AY *et al.*, 2017; LAURITZEN, 1987).

Definição 16 (Tensor de Amari-Chentsov). (AY *et al.*, 2017) Para cada $\mu \in \mathcal{M}_+(\mathcal{X})$, definimos o tensor de Amari-Chentsov T em $\mathcal{M}_+(\mathcal{X})$ como o campo tensorial 3-covariante definido por

$$T_\mu(A, B, C) := T(\mu)(A, B, C) := g_\mu \left(\tilde{\nabla}_A^{(m)} B(\mu) - \tilde{\nabla}_A^{(e)} B(\mu), C(\mu) \right)$$

para todo $\mu \in \mathcal{M}_+(\mathcal{X})$ e para todo $A, B, C \in \mathfrak{X}(\mathcal{M}_+(\mathcal{X}))$.

Sendo $A, B, C \in \mathfrak{X}(\mathcal{M}_+(\mathcal{X}))$ e usando a notação $A(\mu) = (\mu, a_\mu)$, $B(\mu) = (\mu, b_\mu)$, $C(\mu) = (\mu, c_\mu)$, é perceptível que

$$T_\mu(A, B, C) = \sum_{i \in I} \mu_i \frac{a_{\mu i} b_{\mu i} c_{\mu i}}{\mu_i \mu_i \mu_i} \quad (3.12)$$

De fato

$$\begin{aligned} T_\mu(A, B, C) &= \sum_{i \in I} \frac{\left(\tilde{\nabla}_A^{(m)} B(\mu) - \tilde{\nabla}_A^{(e)} B(\mu) \right)_i c_{\mu i}}{\mu_i} \mu_i \\ &= \sum_{i \in I} \frac{\mu_i}{\mu_i} \left(\frac{\partial b_i}{\partial a_\mu}(\mu) - \frac{\partial b_i}{\partial a_\mu}(\mu) + \frac{1}{\mu_i} a_{\mu i} b_{\mu i} \right) \frac{c_{\mu i}}{\mu_i} \\ &= \sum_{i \in I} \mu_i \frac{a_{\mu i} b_{\mu i} c_{\mu i}}{\mu_i \mu_i \mu_i}. \end{aligned}$$

Como é de se imaginar, podemos definir o tensor de Amari-Chentsov em $\mathcal{P}_+(\mathcal{X})$ apenas substituindo $\tilde{\nabla}^{(m)}$ e $\tilde{\nabla}^{(e)}$, respectivamente, por $\nabla^{(m)}$ e $\nabla^{(e)}$. Assim, para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $A, B, C \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$, o tensor de Amari-Chentsov em $\mathcal{P}_+(\mathcal{X})$ é dado por

$$T_\xi(A, B, C) := T(\xi)(A, B, C) := g_\xi \left(\nabla_A^{(m)} B(\xi) - \nabla_A^{(e)} B(\xi), C(\xi) \right).$$

Desse modo, podemos, tal qual fizemos para T em $\mathcal{M}_+(\mathcal{X})$, calcular a expressão para o tensor de Amari-Chentsov em termos de coordenadas. Com efeito,

$$\begin{aligned} T_\xi(A, B, C) &= \sum_{i \in I} \frac{\left(\nabla_A^{(m)} B(\xi) - \nabla_A^{(e)} B(\xi) \right)_i c_{\xi i}}{\xi_i} \xi_i \\ &= \sum_{i \in I} \frac{1}{\xi_i} \left[\frac{\partial b_i}{\partial a_\xi}(\xi) - \frac{\partial b_i}{\partial a_\xi}(\xi) + \frac{1}{\xi_i} a_{\xi i} b_{\xi i} - \xi_i \left(\sum_{j \in I} \frac{1}{\xi_j} a_{\xi j} b_{\xi j} \right) \right] \frac{c_{\xi i}}{\xi_i} \xi_i, \end{aligned}$$

ou seja,

$$T_\xi(A, B, C) = \sum_{i \in I} \frac{a_{\xi i} b_{\xi i} c_{\xi i}}{\xi_i \xi_i \xi_i} \xi_i,$$

para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $A, B, C \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$. Por esse cálculo, podemos observar que o termo adicional que viria da projeção de $\tilde{\nabla}_A^{(e)} B$ não aparece na expressão anterior devido à soma das coordenadas de C ser nula na base $\mathcal{B}_{\mathcal{P}_+(\mathcal{X})}$.

Ademais, como foi apresentado através do Teorema de Chentsov, a métrica de Fisher é caracterizada em $\mathcal{P}_+(\mathcal{X})$ sob sua invariância por morfismos de Markov. Cabe destacar que tal caracterização pode ser estendida ao tensor de Amari-Chentsov em $\mathcal{P}_+(\mathcal{X})$, como expresso no teorema abaixo.

Teorema 3. (AY *et al.*, 2017) Atribua a cada conjunto finito não vazio \mathcal{X} um campo tensorial 3-covariante não trivial $S^{\mathcal{X}}$ em $\mathcal{P}_+(\mathcal{X})$. Se, para cada núcleo de Markov congruente $K : \mathcal{X} \rightarrow \mathcal{X}'$, houver a invariância no sentido

$$S_\xi^{\mathcal{X}}(A, B, C) = S_{K_*(\xi)}^{\mathcal{X}'}(d_\xi K_*(A), d_\xi K_*(B), d_\xi K_*(C)),$$

então existe uma constante $\alpha \neq 0$ tal que $S^{\mathcal{X}} = \alpha T^{\mathcal{X}}$ para todo \mathcal{X} , onde $T^{\mathcal{X}}$ denota o tensor de Amari-Chentsov em $\mathcal{P}_+(\mathcal{X})$.

A demonstração desse teorema é análoga à demonstração de 2, sendo igualmente longa e não acrescentando novas argumentações. Contudo, dado que esse também é um teorema importante no contexto de geometria da informação sobre conjuntos finitos, decidimos apresentar a demonstração dele em A. É importante pontuar que (AY *et al.*, 2017) enuncia esse teorema seguindo as mesmas conclusões de 2, afirmando que $\alpha > 0$. Entretanto, α ser estritamente positiva, no contexto de 2, é uma característica que advém da métrica de Fisher ser positivo definida. Para observar que α pode assumir valores estritamente negativos, considere $S^{\mathcal{X}} := -T^{\mathcal{X}}$, em que todas as hipóteses de 3 são satisfeitas e $\alpha = -1$.

Observando tanto 3.5 quanto 3.12, é fácil notar que podemos estender a definição desses campos tensoriais em $\mathcal{M}_+(\mathcal{X})$ para um campo n -covariante, com $n \geq 1$. Desse modo, para todo $V^{(1)}, V^{(2)}, \dots, V^{(n)} \in \mathfrak{X}(\mathcal{M}_+(\mathcal{X}))$, defina

$$\tau_{\mu}^n \left(V^{(1)}, V^{(2)}, \dots, V^{(n)} \right) := \tau^n(\mu) \left(V^{(1)}, V^{(2)}, \dots, V^{(n)} \right) := \sum_{i \in I} \mu_i \frac{v_{\mu i}^{(1)}}{\mu_i} \frac{v_{\mu i}^{(2)}}{\mu_i} \dots \frac{v_{\mu i}^{(n)}}{\mu_i},$$

com $V^{(i)}(\mu) = \left(\mu, v_{\mu}^{(i)} \right)$ para $i \in I$ (AY *et al.*, 2017). Claramente, $\tau^2 = g$ e $\tau^3 = T$. Além disso, dado um mergulho $p : M \rightarrow \mathcal{M}_+(\mathcal{X})$ de uma variedade diferenciável M em $\mathcal{M}_+(\mathcal{X})$, podemos realizar o *pullback* de τ^n para M através desse mergulho. Sendo $V_1, V_2, \dots, V_n \in \mathfrak{X}(M)$, então, para todo $q \in M$,

$$\begin{aligned} \tau_q^n(V_1, \dots, V_n) &:= p^*(\tau^n)_q(V_1, \dots, V_n) \\ &= \tau_{p(q)}^n \left(\frac{\partial p}{\partial V_1}, \dots, \frac{\partial p}{\partial V_n} \right) \\ &= \sum_{i \in I} \frac{1}{p_i(q)^{n-1}} \frac{\partial p_i}{\partial V_1}(q) \dots \frac{\partial p_i}{\partial V_n}(q) \end{aligned}$$

ou seja,

$$\tau_q^n(V_1, \dots, V_n) = \sum_{i \in I} p_i(q) \frac{\partial \ln p_i}{\partial V_1}(q) \frac{\partial \ln p_i}{\partial V_2}(q) \dots \frac{\partial \ln p_i}{\partial V_n}(q).$$

Considere, agora, o mapa multilinear

$$\begin{aligned} L_{\mathcal{X}}^n : \overbrace{\mathcal{F}(\mathcal{X}) \times \dots \times \mathcal{F}(\mathcal{X})}^n &\longrightarrow \mathbb{R} \\ (f_1, f_2, \dots, f_n) &\longmapsto \sum_{i \in I} f_1^i f_2^i \dots f_n^i \end{aligned}$$

Perceba que ele é natural, sendo uma generalização da métrica usual do \mathbb{R}^n em

$$\mathcal{F}(\mathcal{X})^n := \overbrace{\mathcal{F}(\mathcal{X}) \times \dots \times \mathcal{F}(\mathcal{X})}^n.$$

Considere, também, o mapa

$$\begin{aligned} \pi^{1/n} : \mathcal{M}_+(\mathcal{X}) &\longmapsto \mathcal{F}(x) \\ \mu &\longmapsto n \sum_{i \in I} \mu_i^{\frac{1}{n}} e_i. \end{aligned}$$

Calculando o *pullback* de $L_{\mathcal{X}}^n$ através de $\pi^{1/n}$, obtemos

$$\begin{aligned} L_{\mathcal{X}}^n \left(d_{\mu} \pi^{1/n}(V_1), \dots, d_{\mu} \pi^{1/n}(V_n) \right) &= L_{\mathcal{X}}^n \left(\frac{\partial \pi^{1/n}}{\partial V_1}(\mu), \dots, \frac{\partial \pi^{1/n}}{\partial V_n}(\mu) \right) \\ &= \sum_{i \in I} \left(\mu_i^{\frac{1-n}{n}} v_{1i} \right) \dots \left(\mu_i^{\frac{1-n}{n}} v_{ni} \right) \\ &= \tau_{\mu}^n(V_1, \dots, V_n) \end{aligned}$$

para todo $V_1, V_2, \dots, V_n \in \mathfrak{X}(\mathcal{M}_+(\mathcal{X}))$. Assim, vemos que τ^n surge de forma orgânica através do *pullback* de $L_{\mathcal{X}}^n$ por $\pi^{1/n}$. Como τ^n é uma generalização de g e T , é legítimo questionar se, para todo $n \geq 1$, τ^n é o único campo tensorial n -covariante não trivial em $\mathcal{P}_+(\mathcal{X})$ invariante por morfismos de Markov. A resposta para essa indagação é negativa. Com efeito, para $n = 4$, considere os campos tensoriais abaixo, apresentados em (AY *et al.*, 2017) e definidos em $\mathcal{P}_+(\mathcal{X})$ para todo $V_1, V_2, V_3, V_4 \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$:

$$\begin{aligned} \tau^{\{1,2\}\{3,4\}}(V_1, V_2, V_3, V_4) &= \tau^2(V_1, V_2) \tau^2(V_3, V_4) = g(V_1, V_2) g(V_3, V_4) \\ \tau^{\{1,3\}\{2,4\}}(V_1, V_2, V_3, V_4) &= \tau^2(V_1, V_3) \tau^2(V_2, V_4) = g(V_1, V_3) g(V_2, V_4). \end{aligned}$$

Assuma, agora, que, para cada conjunto finito e não vazio \mathcal{X} , associamos os campos tensoriais $\tau^{\{1,2\}\{3,4\}}$ e $\tau^{\{1,3\}\{2,4\}}$. Devido à invariância da métrica de Fisher por morfismos de Markov, essa invariância estende-se a esses campos tensoriais. De fato, para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $V_1, V_2, V_3, V_4 \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$, temos que

$$\begin{aligned} \left(\tau^{\{1,2\}\{3,4\}} \right)_{\xi}^{\mathcal{X}}(V_1, V_2, V_3, V_4) &= g_{\xi}^{\mathcal{X}}(V_1, V_2) g_{\xi}^{\mathcal{X}}(V_3, V_4) \\ &= g_{K_*(\xi)}^{\mathcal{X}'}(d_{\xi} K_*(V_1), d_{\xi} K_*(V_2)) g_{K_*(\xi)}^{\mathcal{X}'}(d_{\xi} K_*(V_3), d_{\xi} K_*(V_4)) \\ &= \left(\tau^{\{1,2\}\{3,4\}} \right)_{K_*(\xi)}^{\mathcal{X}'}(d_{\xi} K_*(V_1), d_{\xi} K_*(V_2), d_{\xi} K_*(V_3), d_{\xi} K_*(V_4)). \end{aligned}$$

Analogamente, para $\tau^{\{1,3\}\{2,4\}}$, vemos que

$$\left(\tau^{\{1,3\}\{2,4\}} \right)_{\xi}^{\mathcal{X}}(V_1, V_2, V_3, V_4) = \left(\tau^{\{1,3\}\{2,4\}} \right)_{K_*(\xi)}^{\mathcal{X}'}(d_{\xi} K_*(V_1), d_{\xi} K_*(V_2), d_{\xi} K_*(V_3), d_{\xi} K_*(V_4)).$$

Entretanto, não existe constante $\alpha \neq 0$ tal que $\tau^{\{1,2\}\{3,4\}} = \alpha \tau^{\{1,3\}\{2,4\}}$. Realmente, para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $V_1, V_2, V_3, V_4 \in \mathfrak{X}(\mathcal{P}_+(\mathcal{X}))$, considere a sequência de igualdades

$$\begin{aligned} \left(\tau^{\{1,2\}\{3,4\}} \right)_{\xi}^{\mathcal{X}}(V_1, V_2, V_3, V_4) - \alpha \left(\tau^{\{1,3\}\{2,4\}} \right)_{\xi}^{\mathcal{X}}(V_1, V_2, V_3, V_4) &= \\ &= g_{\xi}(V_1, V_2) g_{\xi}(V_3, V_4) - \alpha g_{\xi}(V_1, V_3) g_{\xi}(V_2, V_4) \\ &= g_{\xi}(V_1, g_{\xi}(V_3, V_4) V_2) - g_{\xi}(V_1, \alpha g_{\xi}(V_2, V_4) V_3), \end{aligned}$$

que, dada a arbitrariedade dos campos V_i , $i = 1, 2, 3, 4$, é igual a zero se, e somente se,

$$g_{\xi}(V_3, V_4) V_2 - \alpha g_{\xi}(V_2, V_4) V_3 = 0.$$

Assim, fica claro que não existe $\alpha \neq 0$ que satisfaça a igualdade acima.

Os campos $\tau^{\{1,2\}\{3,4\}}$ e $\tau^{\{1,3\}\{2,4\}}$ são exemplos dos chamados tensores canônicos referentes, neste caso, às partições $P_1 = \{\{1,2\}\{3,4\}\}$ e $P_2 = \{\{1,3\}\{2,4\}\}$, respectivamente. Os tensores canônicos podem ser utilizados para mostrar a inexistência de um teorema de caracterização como o de Chentsov para τ^n em $\mathcal{P}_+(\mathcal{X})$ com $n \geq 4$.

Definição 17 (Tensor canônico). (AY *et al.*, 2017) Seja \mathcal{X} conjunto finito não vazio denotado por $\mathcal{X} = \{x_1, x_2, \dots, x_{m+1}\}$. Dada uma partição $P = \{P_1, P_2, \dots, P_r\}$ de \mathcal{X} , associamos a ela uma bijeção

$$\pi_P: \bigcup_{i \in \{1, \dots, r\}} (\{i\} \times \{1, 2, \dots, n_i\}),$$

onde $n_i = |P_i|$ e tal que $\pi_P(\{i\} \times \{1, 2, \dots, n_i\}) = P_i$, para $i = 1, 2, \dots, r$. Ela está definida a menos de permutações dos elementos de P_i . Definimos o n -tensor canônico de P como o campo tensorial n -covariante τ^P dado, para todo $\mu \in \mathcal{M}_+(\mathcal{X})$ e para todo $V_1, V_2, \dots, V_n \in T_{\mu} \mathcal{M}_+(\mathcal{X})$, como

$$\tau_{\mu}^P(V_1, V_2, \dots, V_n) := \tau^P(\mu)(V_1, V_2, \dots, V_n) := \prod_{i=1}^r \tau_{\mu}^{n_i} \left(V_{\pi^P(i,1)}, V_{\pi^P(i,2)}, \dots, V_{\pi^P(i,n_i)} \right),$$

onde π^P é uma notação sucinta para a composição $\mathcal{I} \circ \pi_P$ e onde

$$\tau_{\mu}^{n_i} \left(V_{\pi^P(i,1)}, \dots, V_{\pi^P(i,n_i)} \right) = \sum_{j \in I} \frac{1}{\mu_j^{n_i-1}} V_{\pi^P(i,1)j} \cdots V_{\pi^P(i,n_i)j}.$$

Munidos dessa definição, estudemos a invariância de τ^n para $n \geq 4$.

Proposição 6. Seja \mathcal{X} um conjunto finito não vazio. Para $n \geq 4$, não existe caracterização de campos tensoriais n -covariante não triviais em $\mathcal{P}_+(\mathcal{X})$ através de invariância por morfismos de Markov induzidos por núcleos de Markov congruentes.

Demonstração. Suponha que n seja par e considere $\text{Part}_2^0(\mathcal{X})$ o conjunto de todas as partições de \mathcal{X} tais que os elementos delas possuem cardinalidade 2. Em outras palavras, todo $P \in \text{Part}_2^0(\mathcal{X})$ é uma coleção de dupletos de \mathcal{X} . Sejam, então, $P_1, P_2 \in \text{Part}_2^0(\mathcal{X})$. Escrevendo $P_1 = (P_1^1, P_2^1, \dots, P_r^1)$ e $P_2 = (P_1^2, P_2^2, \dots, P_r^2)$, suponha que não exista permutação

$$\sigma: \{1, 2, \dots, r\} \rightarrow \{1, 2, \dots, r\}$$

tal que

$$P_2 = \left(P_{\sigma(1)}^1, P_{\sigma(2)}^1, \dots, P_{\sigma(r)}^1 \right).$$

Considere os tensores canônicos dessas partições restritos a $\mathcal{P}_+(\mathcal{X})$. Para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $V_1, V_2, \dots, V_n \in T_{\xi} \mathcal{P}_+(\mathcal{X})$, temos que

$$\begin{aligned} \tau_{\xi}^{P_1}(V_1, V_2, \dots, V_n) &= \prod_{i=1}^r g_{\xi} \left(V_{\pi^{P_1}(i,1)}, V_{\pi^{P_1}(i,2)} \right), \\ \tau_{\xi}^{P_2}(V_1, V_2, \dots, V_n) &= \prod_{j=1}^r g_{\xi} \left(V_{\pi^{P_2}(j,1)}, V_{\pi^{P_2}(j,2)} \right). \end{aligned}$$

Note que a invariância de ambos por morfismos de Markov induzidos por núcleos de Markov congruentes advém da invariância da métrica de Fisher. Além disso, como não existe σ tal qual foi colocada, $\tau^{P_1} \neq \tau^{P_2}$. Suponha por absurdo que exista constante $\alpha \neq 0$ tal que $\tau^{P_1} = \alpha \tau^{P_2}$. Assim, para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $V_1, V_2, \dots, V_n \in T_\xi \mathcal{P}_+(\mathcal{X})$,

$$\tau_\xi^{P_1}(V_1, V_2, \dots, V_n) - \alpha \tau_\xi^{P_2}(V_1, V_2, \dots, V_n) = 0$$

se, e somente se,

$$\prod_{i=1}^r g_\xi(V_{\pi^{P_1}(i,1)}, V_{\pi^{P_1}(i,2)}) - \alpha \prod_{j=1}^r g_\xi(V_{\pi^{P_2}(j,1)}, V_{\pi^{P_2}(j,2)}) = 0. \quad (3.13)$$

Utilizando, novamente, da não existência de σ , existem k, \tilde{k} tais que apenas um dos elementos de P_k^1 é igual a um dos elementos de $P_{\tilde{k}}^2$. Sem perda de generalidade, suponha que $\pi^{P_1}(k, 1) = \pi^{P_2}(\tilde{k}, 1)$ e $\pi^{P_1}(k, 2) \neq \pi^{P_2}(\tilde{k}, 2)$. Assim, 3.13 é verdadeira se, e somente se,

$$g_\xi(V_{\pi^{P_1}(k,1)}, V_{\pi^{P_1}(k,2)}) \prod_{i \neq k}^r g_\xi(V_{\pi^{P_1}(i,1)}, V_{\pi^{P_1}(i,2)}) - \alpha g_\xi(V_{\pi^{P_2}(\tilde{k},1)}, V_{\pi^{P_2}(\tilde{k},2)}) \prod_{j \neq \tilde{k}}^r g_\xi(V_{\pi^{P_2}(j,1)}, V_{\pi^{P_2}(j,2)}) = 0$$

que, por sua vez, é válida se, e somente se,

$$\prod_{i \neq k}^r g_\xi(V_{\pi^{P_1}(i,1)}, V_{\pi^{P_1}(i,2)}) V_{\pi^{P_1}(k,2)} - \alpha \prod_{j \neq \tilde{k}}^r g_\xi(V_{\pi^{P_2}(j,1)}, V_{\pi^{P_2}(j,2)}) V_{\pi^{P_2}(\tilde{k},2)} = 0.$$

Como $\pi^{P_1}(k, 2) \neq \pi^{P_2}(\tilde{k}, 2)$, a igualdade acima apenas é verdadeira se $\{V_{\pi^{P_1}(k,2)}, V_{\pi^{P_2}(\tilde{k},2)}\}$ for um conjunto linearmente dependente, o que é um absurdo, pois $V_1, V_2, \dots, V_n \in T_\xi \mathcal{P}_+(\mathcal{X})$ foram tomados arbitrariamente. Logo, para n par, não existe uma caracterização de τ^n via invariância por morfismos de Markov induzidos por núcleos de Markov congruentes. Já se n for ímpar, seja $\text{Part}_2^3(\mathcal{X})$ o conjunto das partições de \mathcal{X} que contém $r := \frac{n-3}{2}$ dupletos e um conjunto de cardinalidade 3. Assim, sejam $P_1, P_2 \in \text{Part}_2^3(\mathcal{X})$. Como antes, escrevendo $P_1 = (P_1^1, P_2^1, \dots, P_r^1)$ e $P_2 = (P_1^2, P_2^2, \dots, P_r^2)$, suponha que não exista permutação

$$\sigma : \{1, 2, \dots, r+1\} \rightarrow \{1, 2, \dots, r+1\}$$

tal que

$$P_2 = (P_{\sigma(1)}^1, P_{\sigma(2)}^1, \dots, P_{\sigma(r+1)}^1).$$

Além disso, suponha, sem perda de generalidade, que P_l^1 e P_l^2 são os conjuntos de dois elementos de P_1 e P_2 , respectivamente. Considere, agora, os tensores canônicos dessas partições. Para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $V_1, V_2, \dots, V_n \in T_\xi \mathcal{P}_+(\mathcal{X})$, temos que

$$\begin{aligned} \tau_\xi^{P_1}(V_1, \dots, V_n) &= T_\xi(V_{\pi^{P_1}(l,1)}, V_{\pi^{P_1}(l,2)}, V_{\pi^{P_1}(l,3)}) \prod_{i=1}^r g_\xi(V_{\pi^{P_1}(i,1)}, V_{\pi^{P_1}(i,2)}) \\ \tau_\xi^{P_2}(V_1, \dots, V_n) &= T_\xi(V_{\pi^{P_2}(\tilde{l},1)}, V_{\pi^{P_2}(\tilde{l},2)}, V_{\pi^{P_2}(\tilde{l},3)}) \prod_{j=1}^r g_\xi(V_{\pi^{P_2}(j,1)}, V_{\pi^{P_2}(j,2)}). \end{aligned}$$

A invariância desses campos por morfismos de Markov induzidos por núcleos de Markov congruentes advém da invariância de g e de T . A partir deste ponto, a argumentação segue análoga ao caso de n par. Como não existe σ tal como foi colocada, $\tau_\xi^{P_1} \neq \tau_\xi^{P_2}$. Suponha, por absurdo, que exista constante $\alpha \neq 0$ tal que $\tau_\xi^{P_1} = \alpha \tau_\xi^{P_2}$. Novamente, da não existência da permutação σ , existem k, \tilde{k} tais que apenas um elemento de P_k^1 é igual a um dos elementos de $P_{\tilde{k}}^2$. Sem perda de generalidade, vamos supor que $\pi^{P_1}(k, 1) = \pi^{P_2}(\tilde{k}, 1)$ e $\pi^{P_1}(k, 2) \neq \pi^{P_2}(\tilde{k}, 2)$. Desse modo, para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $V_1, V_2, \dots, V_n \in T_\xi \mathcal{P}_+(\mathcal{X})$,

$$\tau_\xi^{P_1}(V_1, \dots, V_n) - \alpha \tau_\xi^{P_2}(V_1, \dots, V_n) = 0$$

se, e somente se,

$$\begin{aligned} T_\xi \left(V_{\pi^{P_1}(l,1)}, V_{\pi^{P_1}(l,2)}, V_{\pi^{P_1}(l,3)} \right) \prod_{i \neq k}^r g_\xi \left(V_{\pi^{P_1}(i,1)}, V_{\pi^{P_1}(i,2)} \right) V_{\pi^{P_1}(k,2)} \\ - \alpha T_\xi \left(V_{\pi^{P_2}(\tilde{l},1)}, V_{\pi^{P_2}(\tilde{l},2)}, V_{\pi^{P_2}(\tilde{l},3)} \right) \prod_{j \neq \tilde{k}}^r g_\xi \left(V_{\pi^{P_2}(j,1)}, V_{\pi^{P_2}(j,2)} \right) V_{\pi^{P_2}(\tilde{k},2)} = 0. \end{aligned}$$

Contudo, essa igualdade apenas é verdadeira se $\left\{ V_{\pi^{P_1}(k,2)}, V_{\pi^{P_2}(\tilde{k},2)} \right\}$ for um conjunto linearmente dependente, o que é um absurdo. \square

Um fato interessante é que τ^1 é nulo em $\mathcal{P}_+(\mathcal{X})$, uma vez que, para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $V \in T_\xi \mathcal{P}_+(\mathcal{X})$, $\tau_\xi^1(V) = \sum_{i=1}^{m+1} v_i = 0$. Devido a isso, na argumentação acima, não utilizamos partições que contém singletos, já que seus tensores canônicos são triviais.

3.5 As α -conexões

Retornando ao estudo das conexões tanto em $\mathcal{M}_+(\mathcal{X})$ quanto em $\mathcal{P}_+(\mathcal{X})$, as conexões mistura e exponencial nesses espaços podem ser entendidas como casos destacados dentro de família de conexões lineares: as α -conexões. Em contrapartida, elas também permitem a definição dessa família, como segue abaixo.

Definição 18 (α -conexões). (AY *et al.*, 2017) Seja $\alpha \in [-1, 1]$. Definimos a α -conexão $\tilde{\nabla}^{(\alpha)}$ em $\mathcal{M}_+(\mathcal{X})$ como a combinação convexa²⁰

$$\tilde{\nabla}^{(\alpha)} := \frac{1-\alpha}{2} \tilde{\nabla}^{(m)} + \frac{1+\alpha}{2} \tilde{\nabla}^{(e)} = \tilde{\nabla}^{(m)} + \frac{1+\alpha}{2} \left(\tilde{\nabla}^{(e)} - \tilde{\nabla}^{(m)} \right).$$

É perceptível que, como combinação convexa de conexões lineares em $\mathcal{M}_+(\mathcal{X})$, $\tilde{\nabla}^{(\alpha)}$ também é uma conexão linear nessa variedade. Além disso, é importante pontuar que o tensor de Amari-Chentsov pode ser interpretado, grosso modo, como uma medida de quanto a métrica

²⁰ Por combinação convexa de $\tilde{\nabla}^{(m)}$ e $\tilde{\nabla}^{(e)}$ entende-se uma combinação dessas conexões tal que os coeficientes são não negativos e têm soma unitária.

de Fisher distingue a α -conexão da conexão mistura. Em outras palavras, dados $A, B, C \in \mathfrak{X}(\mathcal{M}_+(\mathcal{X}))$, podemos escrever

$$\begin{aligned} g\left(\tilde{\nabla}_A^{(\alpha)} B, C\right) &= g\left(\tilde{\nabla}_A^{(m)} B, C\right) + \frac{1+\alpha}{2} g\left(\tilde{\nabla}_A^{(e)} B - \tilde{\nabla}_A^{(m)} B, C\right) \\ &= g\left(\tilde{\nabla}_A^{(m)} B, C\right) - \frac{1+\alpha}{2} T(A, B, C), \end{aligned}$$

onde, em termos das coordenadas canônicas e para todo $\mu \in \mathcal{M}_+(\mathcal{X})$, $\nabla_A^{(\alpha)} B$ é dado por

$$\tilde{\nabla}_A^{(\alpha)} B(\mu) = \left(\mu, \sum_{i \in I} \left(\frac{\partial b_i}{\partial a_\mu}(\mu) - \frac{1+\alpha}{2} \frac{a_{\mu i} b_{\mu i}}{\mu_i} \right) \delta^i \right).$$

Ademais, sendo $\tilde{\nabla}^{(\alpha)}$ uma conexão em $\mathcal{M}_+(\mathcal{X})$, podemos calcular as geodésicas máximas com respeito a ela, chamadas de α -geodésicas²¹.

Proposição 7. A α -geodésica maximal em $\mathcal{M}_+(\mathcal{X})$ com ponto inicial $\mu \in \mathcal{M}_+(\mathcal{X})$ e velocidade inicial $a \in T_\mu \mathcal{M}_+(\mathcal{X})$ é dada por

$$\begin{aligned} \tilde{\gamma}^{(\alpha)} : (t^-, t^+) &\longrightarrow \mathcal{M}_+(\mathcal{X}) \\ t &\longmapsto \left(\mu^{\frac{1-\alpha}{2}} + t \frac{(1-\alpha)}{2} \mu^{-\frac{1+\alpha}{2}} a \right)^{\frac{2}{1-\alpha}}, \end{aligned}$$

onde

$$t^- := -\min \left\{ \frac{2\mu_i}{(1-\alpha)a_i} \mid i \in I, a_i > 0 \right\}, \quad t^+ := \min \left\{ \frac{2\mu_i}{(1-\alpha)|a_i|} \mid i \in I, a_i < 0 \right\}.$$

Demonstração. Considere $\mu \in \mathcal{M}_+(\mathcal{X})$. Então, os símbolos de Christoffel de $\tilde{\nabla}^{(\alpha)}$ são

$$\begin{aligned} \tilde{\Gamma}_{ij}^{(\alpha)k}(\mu) &= g_\mu \left(\tilde{\nabla}_i^{(\alpha)} \partial_j, \partial_k \right) \mu_k \\ &= \mu_k \left(g_\mu \left(\tilde{\nabla}_i^{(m)} \partial_j, \partial_k \right) - \frac{1+\alpha}{2} T_\mu(\partial_i, \partial_j, \partial_k) \right) \\ &= -\frac{1+\alpha}{2} \mu_k T_\mu(\partial_i, \partial_j, \partial_k) \end{aligned}$$

ou seja,

$$\tilde{\Gamma}_{ij}^{(\alpha)k}(\mu) = -\frac{1+\alpha}{2\mu_k} \delta_{ij} \delta_{ik}.$$

Assim, a equação da geodésica escreve-se como

$$\ddot{\gamma} - \frac{1+\alpha}{2} \frac{\dot{\gamma}}{\gamma} = 0.$$

²¹ Tal qual foi feito para as m - e e -conexões, indicaremos as grandezas referentes a $\tilde{\nabla}^{(\alpha)}$ adicionando a elas o prefixo “ α ” ou o superíndice “ (α) ”.

Afirmamos que, impondo as condições iniciais $\gamma(0) = \mu \in \mathcal{M}_+(\mathcal{X})$ e $\dot{\gamma}(0) = a \in T_\mu \mathcal{M}_+(\mathcal{X})$, a solução da equação anterior é

$$\tilde{\gamma}^{(\alpha)}(t) = \left(\mu^{\frac{1-\alpha}{2}} + t \frac{1-\alpha}{2} \mu^{-\frac{1+\alpha}{2}} a \right)^{\frac{2}{1-\alpha}}.$$

De fato, primeiramente, perceba que $\tilde{\gamma}^{(\alpha)}(0) = \mu$. Além disso, note que

$$\dot{\tilde{\gamma}}^{(\alpha)}(t) = \frac{a}{\mu^{\frac{\alpha+1}{2}}} \cdot \left(\frac{a(1-\alpha)t}{2\mu^{\frac{\alpha+1}{2}}} + \mu^{\frac{1-\alpha}{2}} \right)^{\frac{1+\alpha}{1-\alpha}}$$

e que

$$\ddot{\tilde{\gamma}}^{(\alpha)}(t) = \frac{a^2(1+\alpha)}{2\mu^{\alpha+1}} \left(\mu^{\frac{1-\alpha}{2}} + \frac{a(1-\alpha)t}{2\mu^{\frac{\alpha+1}{2}}} \right)^{\frac{2\alpha}{1-\alpha}},$$

onde $a^2 = \sum_{i \in I} a_i^2$. Dessa maneira,

$$\dot{\tilde{\gamma}}^{(\alpha)}(0) = \frac{a}{\mu^{\frac{\alpha+1}{2}}} \mu^{\frac{1+\alpha}{2}} = a$$

e

$$\begin{aligned} \ddot{\tilde{\gamma}}^{(\alpha)} - \frac{1+\alpha}{2} \frac{\dot{\tilde{\gamma}}^{(\alpha)2}}{\tilde{\gamma}^{(\alpha)}} &= \frac{a^2(1+\alpha)}{2\mu^{\alpha+1}} \left(\mu^{\frac{1-\alpha}{2}} + \frac{a(1-\alpha)t}{2\mu^{\frac{\alpha+1}{2}}} \right)^{\frac{2\alpha}{1-\alpha}} \\ &\quad - \frac{1+\alpha}{2} \frac{a^2}{\mu^{\alpha+1}} \left(\mu^{\frac{1-\alpha}{2}} + \frac{a(1-\alpha)t}{2\mu^{\frac{\alpha+1}{2}}} \right)^{\frac{2(1+\alpha)}{1-\alpha}} \left(\mu^{\frac{1-\alpha}{2}} + \frac{a(1-\alpha)t}{2\mu^{\frac{\alpha+1}{2}}} \right)^{\frac{-2}{1-\alpha}} \\ &= 0. \end{aligned}$$

Logo, $\tilde{\gamma}^{(\alpha)}$ é α -geodésica satisfazendo as condições iniciais propostas. Perceba que $\tilde{\gamma}^{(\alpha)}$ está bem definida para²²

$$\mu^{\frac{1-\alpha}{2}} + t \frac{1-\alpha}{2} \mu^{-\frac{1+\alpha}{2}} a > 0.$$

Assim, para $i \in I$, estudemos a desigualdade acima dividindo esse estudo em dois casos. Primeiro, suponha que $a_i > 0$. Então,

$$\mu_i^{\frac{1-\alpha}{2}} + t \frac{(1-\alpha)}{2} \mu_i^{-\frac{1+\alpha}{2}} a_i > 0 \Leftrightarrow 2\mu_i + t(1-\alpha)a_i > 0 \Leftrightarrow t > \frac{2\mu_i}{(1-\alpha)a_i}.$$

Já, se $a_i < 0$,

$$\mu_i^{\frac{1-\alpha}{2}} + t \frac{(1-\alpha)}{2} \mu_i^{-\frac{1+\alpha}{2}} a_i > 0 \Leftrightarrow 2\mu_i - t(1-\alpha)(-a_i) > 0 \Leftrightarrow t < \frac{2\mu_i}{(1-\alpha)(-a_i)}.$$

Desse modo, concluímos que $\gamma^{(\alpha)}$ está definida em (t^+, t^-) , onde

$$t^- := -\min \left\{ \frac{2\mu_i}{(1-\alpha)a_i} \mid i \in I, a_i > 0 \right\}, \quad t^+ := \min \left\{ \frac{2\mu_i}{(1-\alpha)|a_i|} \mid i \in I, a_i < 0 \right\}.$$

□

²² Isso expressa o fato do polinômios x^β , com $\beta \in \mathbb{R}$, ser definido, usualmente, como $x^\beta := e^{\beta \ln(x)}$.

Por conseguinte, com $\widetilde{\nabla}^{(\alpha)}$, $\mathcal{M}_+(\mathcal{X})$ não é geodesicamente completa. Para verificar tal fato, fixe $\mu \in \mathcal{M}_+(\mathcal{X})$ e $a \in T_\mu \mathcal{M}_+(\mathcal{X})$ e considere $\alpha = -1$. Nesse caso, $\widetilde{\gamma}^{(\alpha)} = \mu + ta$, que consiste em uma reta passando por μ com direção a e que, claramente, não fica contida em $\mathcal{M}_+(\mathcal{X})$. Analisemos, agora, o que ocorre quanto α tende a 1 pela esquerda, uma vez que $\alpha = 1$ na expressão 6 produz uma indeterminação. Para tal, observe que

$$\lim_{\alpha \rightarrow 1^-} \frac{\left(\frac{at(1-\alpha)}{2\mu^{\frac{\alpha+1}{2}}} + \mu^{\frac{1-\alpha}{2}} - 1 \right)}{1-\alpha} = \frac{\mu \ln \mu + at}{2\mu}$$

Fazendo a mudança de variável

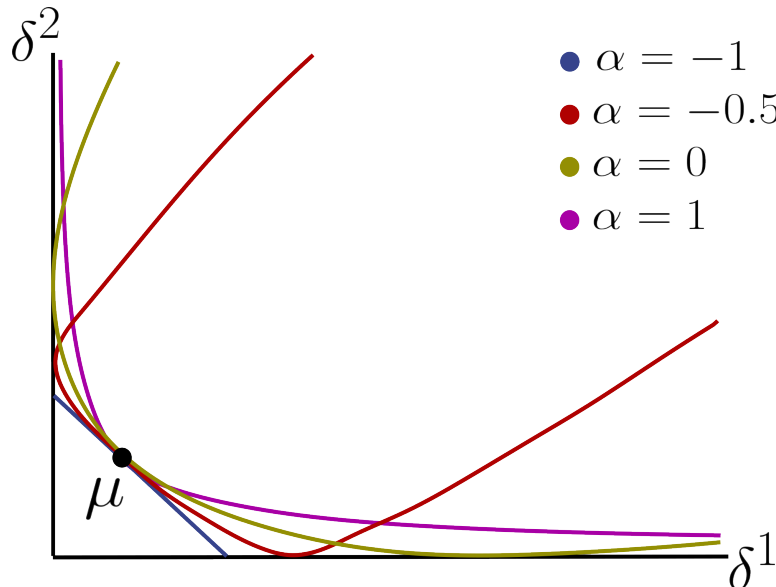
$$\beta = \left(\mu^{\frac{1-\alpha}{2}} + t \frac{(1-\alpha)}{2} \mu^{-\frac{1+\alpha}{2}} a - 1 \right)^{-1},$$

temos, quando $\alpha \rightarrow 1^-$, que $\beta \rightarrow +\infty$. Logo,

$$\begin{aligned} \lim_{\alpha \rightarrow 1^-} \left(\mu^{\frac{1-\alpha}{2}} + t \frac{(1-\alpha)}{2} \mu^{-\frac{1+\alpha}{2}} a \right)^{\frac{2}{1-\alpha}} &= \left[\lim_{\beta \rightarrow +\infty} \left(1 + \frac{1}{\beta} \right)^\beta \right]^{\frac{\mu \ln \mu + at}{\mu}} \\ &= e^{\frac{\mu \ln \mu + at}{\mu}} \\ &= \mu e^{\frac{at}{\mu}}, \end{aligned}$$

ou seja, quando α tende a 1 pela esquerda, a α -geodésica $\gamma^{(\alpha)}(t)$ se aproxima cada vez mais da curva dada pela exponencial $\mu e^{at/\mu}$. Essa é uma das justificativas para chamar a conexão $\nabla^{(e)}$ de conexão exponencial, uma vez que $\nabla^{(1)} = \nabla^{(e)}$. Discutido isso, calculemos o mapa exponencial relativo à α -conexão.

Figura 10 – Representação gráfica de α -geodésicas com $\dim \mathcal{M}_+(\mathcal{X}) = 2$ e $\alpha = -1, -\frac{1}{2}, 0$ e 1. Para que se possa identificar as proporções, foram tomados $\mu = (1, 2)$ e $a = (-4, 5)$



Fonte: Elaborada pelo autor.

Corolário 3. O mapa exponencial $\widetilde{\text{exp}}^{(\alpha)}$ respectivo a $\widetilde{\nabla}^{(\alpha)}$ é dado por

$$\begin{aligned} \widetilde{\text{exp}}^{(\alpha)} : \widetilde{W} &\longrightarrow \mathcal{M}_+(\mathcal{X}) \\ (\mu, a) &\longmapsto \left(\mu^{\frac{1-\alpha}{2}} + \frac{(1-\alpha)}{2} \mu^{-\frac{1+\alpha}{2}} a \right)^{\frac{2}{1-\alpha}}, \end{aligned}$$

com

$$\widetilde{W}^{(\alpha)} := \left\{ (\mu, a) \in T\mathcal{M}_+(\mathcal{X}) \mid \mu \in \mathcal{M}_+(\mathcal{X}), a \in \mathcal{S}(\mathcal{X}), \min \left\{ \frac{2\mu_i}{(1-\alpha)|a_i|} \mid i \in I \right\} > 1 \right\}.$$

Demonstração. Dado $(\mu, a) \in T\mathcal{M}_+(\mathcal{X})$, observamos que, devido aos limites de definição t^- e t^+ de $\widetilde{\gamma}^{(\alpha)}$, $\widetilde{\text{exp}}^{(\alpha)}$ é definido em $\widetilde{W}^{(\alpha)} \subset T\mathcal{M}_+(\mathcal{X})$, com $\widetilde{W}^{(\alpha)}$ conforme o enunciado. Assim, para cada $(\mu, a) \in \widetilde{W}^{(\alpha)}$, considere a α -geodésica $\widetilde{\gamma}^{(\alpha)}$ tal que $\widetilde{\gamma}^{(\alpha)}(0) = \mu$ e $\dot{\widetilde{\gamma}}^{(\alpha)} = a$. Tomando $t = 1$, obtemos

$$\widetilde{\text{exp}}^{(\alpha)}(\mu, a) = \widetilde{\gamma}^{(\alpha)}(1) = \left(\mu^{\frac{1-\alpha}{2}} + \frac{(1-\alpha)}{2} \mu^{-\frac{1+\alpha}{2}} a \right)^{\frac{2}{1-\alpha}}.$$

□

Agora, com o intuito de definirmos as α -conexões $\nabla^{(\alpha)}$ em $\mathcal{P}_+(\mathcal{X})$, projetemos $\widetilde{\nabla}^{(\alpha)}$ em $\mathcal{P}_+(\mathcal{X})$, isto é, para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ e para todo $A, B \in T_\xi \mathcal{P}_+(\mathcal{X})$,

$$\nabla_A^{(\alpha)} B(\xi) := \Pi_\xi^\top \left(\widetilde{\nabla}_A^{(\alpha)} B(\xi) \right),$$

onde Π_ξ^\top é dada por 4. Perceba que, para cada $\xi \in \mathcal{P}_+(\mathcal{X})$, Π_ξ^\top é linear. Usando esse fato, observamos que

$$\begin{aligned} \nabla_A^{(\alpha)} B(\xi) &= \Pi_\xi^\perp \left(\widetilde{\nabla}_A^{(m)} B(\xi) + \frac{1+\alpha}{2} \left(\widetilde{\nabla}_A^{(e)} B(\xi) - \widetilde{\nabla}_A^{(m)} B(\xi) \right) \right) \\ &= \sum_{i \in I} \left\{ \frac{\partial b_i}{\partial a_\xi}(\xi) + \frac{1+\alpha}{2} \left[\frac{\partial b_i}{\partial a_\xi}(\xi) - \frac{1}{\xi_i} a_{\xi_i} b_{\xi_i} + \xi_i \left(\sum_{j \in I} \frac{1}{\xi_j} a_{\xi_j} b_{\xi_j} \right) - \frac{\partial b_i}{\partial a_\xi}(\xi) \right] \right\} \delta^i \end{aligned}$$

ou seja,

$$\nabla_A^{(\alpha)} B(\xi) = \left(\xi, \sum_{i \in I} \left(\frac{\partial b_i}{\partial a_\xi}(\xi) - \frac{1+\alpha}{2} \left[\frac{a_{\xi_i} b_{\xi_i}}{\xi_i} - \xi_i \sum_{j \in I} \frac{a_{\xi_j} b_{\xi_j}}{\xi_j} \right] \right) \right) \delta^i.$$

Assim, escrevendo a equação das α -geodésicas em $\mathcal{P}_+(\mathcal{X})$, obtemos

$$\nabla_{\dot{\gamma}(t)}^{(\alpha)} \dot{\gamma}(t) = \left(\gamma(t), \sum_{i \in I} \left(\dot{\gamma}_i(\gamma(t)) - \frac{1+\alpha}{2} \left[\frac{\dot{\gamma}^2(\gamma(t))}{\gamma_i(t)} - \gamma_i(t) \sum_{j \in I} \frac{\dot{\gamma}_j^2(\gamma(t))}{\gamma_j(t)} \right] \right) \right) \delta^i = (\gamma(t), 0)$$

ou, mais diretamente,

$$\ddot{\gamma} - \frac{1+\alpha}{2} \left(\frac{\dot{\gamma}^2}{\gamma} - \gamma \sum_{i \in I} \frac{\dot{\gamma}_i^2}{\gamma_i} \right) = 0, \quad (3.14)$$

cuja solução $\gamma^{(\alpha)}$ é única sob as condições iniciais

$$\gamma^{(\alpha)}(0) = \xi \in \mathcal{P}_+(\mathcal{X}) \quad \text{e} \quad \dot{\gamma}^{(\alpha)} = a \in T_\xi \mathcal{P}_+(\mathcal{X}).$$

Analisando 3.14, é perceptível que sua resolução apresenta elevada dificuldade, uma vez que 3.14 consiste em um sistema de equações diferenciais não lineares de segunda ordem com dependências não lineares entre si e com $\alpha \in [-1, 1]$ arbitrário. Os autores de (AY *et al.*, 2017), com base no trabalho (MOROZOVA; CHENTSOV, 1991), propõem o *ansatz* de solução

$$\gamma^{(\alpha)}(t) = \sum_{i \in I} \frac{\xi_i \left(1 + \tau_{\xi,a}(t) \frac{1-\alpha}{2} \frac{a_i}{\xi_i}\right)^{\frac{2}{1-\alpha}}}{\sum_{j \in I} \xi_j \left(1 + \tau_{\xi,a}(t) \frac{1-\alpha}{2} \frac{a_j}{\xi_j}\right)^{\frac{2}{1-\alpha}}} \delta^i$$

em termos de uma reparametrização $\tau_{\xi,a}(t)$ desconhecida, mas provida das propriedades

$$\tau_{\xi,a}(0) = 0 \quad \text{e} \quad \dot{\tau}_{\xi,a}(0) = 1.$$

No caso em que $\alpha = -1$, $\gamma^{(\alpha)} = \gamma^{(m)}$ e, conseqüentemente, $\tau_{\xi,a}(t) = t$. Já, quando α tende a 1 pela esquerda, podemos observar que

$$\begin{aligned} \lim_{\alpha \rightarrow 1^-} \xi_i \left(1 + \tau_{\xi,a}(t) \frac{1-\alpha}{2} \frac{a_i}{\xi_i}\right)^{\frac{2}{1-\alpha}} &= \xi_i \exp\left(\frac{a_i \tau_{\xi,a}(t)}{\xi_i}\right), \\ \lim_{\alpha \rightarrow 1^-} \sum_{j \in I} \xi_j \left(1 + \tau_{\xi,a}(t) \frac{1-\alpha}{2} \frac{a_j}{\xi_j}\right)^{\frac{2}{1-\alpha}} &= \sum_{j \in I} \xi_j \exp\left(\frac{a_j \tau_{\xi,a}(t)}{\xi_j}\right), \end{aligned}$$

de maneira que

$$\lim_{\alpha \rightarrow 1^-} \gamma^{(\alpha)}(t) = \sum_{i \in I} \frac{\xi_i \exp\left(\frac{a_i \tau_{\xi,a}(t)}{\xi_i}\right)}{\sum_{j \in I} \xi_j \exp\left(\frac{a_j \tau_{\xi,a}(t)}{\xi_j}\right)} \delta^i.$$

Como $\gamma^{(\alpha)} \rightarrow \gamma^{(e)}$ quando $\alpha \rightarrow 1^-$, comparando a expressão anterior com a fornecida por 5 para as e -geodésicas, observamos que, neste segundo caso limite, $\tau_{\xi,a}$ iguala-se a²³ t . Nesses dois casos destacados, a expressão para $\tau_{\xi,a}(t)$ é facilmente identificada, uma vez que são conhecidas as m - e e -geodésicas. Sem tal condição, essa tarefa torna-se mais complexa.

Ademais e por fim, as α -conexões, compreendidas como generalizações das conexões mistura e exponencial, podem ser entendidas, assim como a métrica de Fisher e o tensor de Amari-Chentsov, como “estruturas”²⁴ que fornecem um caráter dualístico entre geometria e estatística em um contexto finito. Elas foram apresentadas como grandezas no espaço das medidas de probabilidade sobre um conjunto finito não vazio \mathcal{X} induzidas do espaço de medidas finitas sobre esse mesmo conjunto. Agora, estudemos como essas estruturas podem ser introduzidas de maneira intrínseca em uma variedade riemanniana, conjuntura que se classifica como o estudo das variedades estatísticas.

²³ Rigorosamente, como o que foi feito neste caso para justificar a expressão para $\tau_{\xi,a}(t)$ foi a realização de um limite, a afirmação mais precisa é que $\tau_{\xi,a}(t)$ tende assintoticamente a t quando α tende a 1 pela esquerda.

²⁴ Essa escolha de nomenclatura ficará clara no início capítulo seguinte.

VARIEDADES ESTATÍSTICAS E VARIEDADES DUALÍSTICAS

4.1 Estruturas dualística e estatística em uma variedade diferenciável

Em 2, apresentamos os modelos estatísticos como objetos que, primeiramente, são usados dentro da estatística mas que, da forma que foram postos, possuem um potencial geométrico intrínseco. Neste capítulo, seguindo os ideários de Amari e Nagaoka e as noções desenvolvidas por Lauritzen, iniciaremos os estudos já com uma variedade riemanniana, munido-a com estruturas que refletem os aspectos estatísticos desenvolvidos em 2. Para tal, comecemos definindo o que são conexões duais.

Definição 19 (Conexões duais). (NIELSEN, 2020) Seja (M, g) uma variedade riemanniana munida de uma conexão linear ∇ . Uma conexão linear ∇^* nessa variedade é dita ser dual a ∇ com respeito a g se¹, para todo $A, B, C \in \mathfrak{X}(M)$,

$$Cg(A, B) = g(\nabla_C A, B) + g(A, \nabla_C^* B). \quad (4.1)$$

A definição anterior pode ser interpretada, então, como um enfraquecimento da noção de compatibilidade com a métrica, usualmente atrelada à conexão de Levi-Civita. Essa última pode ser entendida, por 4.1, como uma conexão autodual, ou seja, uma conexão tal que $\nabla = \nabla^*$. Além disso, observando 4.1, é perceptível que tomar a conexão dual² é um processo involutivo, ou seja, $(\nabla^*)^* = \nabla$ para todo par de conexões lineares duais entre si (NIELSEN, 2020).

¹ Em algumas literaturas, como em (NIELSEN, 2020), essas conexões podem ser chamadas de conjugadas uma a outra com respeito a g . Além disso, quando ficar subentendida a métrica riemanniana g , omitiremos a designação “com respeito a g ”.

² Dadas duas conexões duais ∇, ∇^* com respeito a uma métrica riemanniana, tomar a dual de uma delas, digamos ∇ , é um processo chamado de dualização ou conjugação de ∇ .

Apresentada, então, a noção de dualidade entre conexões, usemos-na para definir variedades dualísticas.

Definição 20 (Variedade dualística). (AY *et al.*, 2017) Seja M uma variedade diferenciável. Uma estrutura dualística em M é uma tripla ordenada (g, ∇, ∇^*) , onde³ g é uma métrica riemanniana em M e ∇, ∇^* são duas conexões duais com respeito a g . A quádrupla ordenada (M, g, ∇, ∇^*) é chamada de variedade dualística.

Assim, conhecido já o fato de que as α -conexões são livres de torção, é esperado que estruturas dualísticas livres de torção⁴ desempenham uma função destacada. De fato, elas são utilizadas para definir o que Steffen Lauritzen chamou de tensor de assimetria (LAURITZEN, 1987).

Definição 21 (Tensor 3-simétrico). (AY *et al.*, 2017) Sejam ∇, ∇^* conexões livres de torção e duais com respeito a uma métrica riemanniana g em uma variedade diferenciável M . Então, o campo tensorial 3-covariante

$$T := g(\nabla^* - \nabla, \cdot) \quad (4.2)$$

em⁵ M é chamado um tensor 3-simétrico da estrutura dualística (g, ∇, ∇^*) .

Essa definição é motivada pelo conceito de variedade estatística, de maneira que uma estrutura dualística em uma variedade diferenciável induz uma estrutura estatística nela.

Definição 22 (Variedade estatística). (AY *et al.*, 2017) Uma estrutura estatística em uma variedade diferenciável M consiste em uma métrica riemanniana g e em um campo tensorial covariante T de *rank* 3 e simétrico em todos os seus argumentos. Uma variedade estatística é uma variedade diferenciável M equipada com uma estrutura estatística⁶.

Ademais, a expressão 4.2 em um sistema de coordenadas de M escreve-se como

$$T_{ijk} = \Gamma_{ijk}^* - \Gamma_{ijk}, \quad (4.3)$$

onde Γ_{ijk}^* e Γ_{ijk} correspondem, respectivamente, aos símbolos de Christoffel de ∇^* e ∇ com um índice abaixado. De fato,

$$T_{ijk} = g(\nabla_i^* \partial_j - \nabla_i \partial_j, \partial_k) = \sum_{l \in J} (\Gamma_{ij}^{*l} - \Gamma_{ij}^l) g_{lk} = \Gamma_{ijk}^* - \Gamma_{ijk}.$$

Mostremos, agora, que, para além do nome, T é simétrico em seus argumentos.

³ Neste capítulo, tanto g quanto T não indicam, necessariamente, a métrica de Fisher e o tensor de Amari-Chentsov, nessa ordem.

⁴ Por “estruturas dualísticas livres de torção” entende-se estruturas dualísticas em que ∇ e ∇^* são livres de torção.

⁵ Usualmente, a expressão 4.2 é denotada por $T := \nabla^* - \nabla$.

⁶ Por completude, podemos pontuar que duas estruturas dualísticas são equivalentes em uma variedade estatística quando elas se diferem por um difeomorfismo.

Proposição 8. (AY *et al.*, 2017) Um tensor 3-simétrico T de uma estrutura dualística em uma variedade diferenciável M é simétrico em seus três índices.

Demonstração. Como ∇ e ∇^* são livres de torção, pela 4.3, observamos que T_{ijk} é simétrico nos índices i e j . Assim, precisamos mostrar apenas que ele é simétrico em j e k . Por um lado, como ∇ e ∇^* são duais um em relação ao outro, então, para todo $A, B, C \in \mathfrak{X}(M)$, temos que

$$Ag(B, C) = g(\nabla_A B, C) + g(B, \nabla_A^* C). \quad (4.4)$$

Por outro lado, sabemos que

$$Ag(B, C) = (\nabla_A g)(B, C) + g(\nabla_A B, C) + g(B, \nabla_A C). \quad (4.5)$$

Comparando 4.4 e 4.5, obtemos que

$$(\nabla_A g)(B, C) = g(B, (\nabla_A^* - \nabla_A)C).$$

Escrevendo a igualdade acima em termos de um sistema de coordenadas arbitrário em M , temos a sequência de igualdades

$$\begin{aligned} (\nabla_i g)(\partial_j, \partial_k) &= (\nabla_i g)_{jk} \\ &= g(\partial_j, (\nabla_i^* - \nabla_i)\partial_k) \\ &= g\left(\partial_j, \sum_{l \in J} (\Gamma_{ik}^{*l} - \Gamma_{ik}^l)\partial_l\right) \\ &= \sum_{l \in J} g_{jl} (\Gamma_{ik}^{*l} - \Gamma_{ik}^l) \\ &= T_{ikj}. \end{aligned}$$

Como $(\nabla_i g)(\partial_j, \partial_k) = (\nabla_i g)(\partial_k, \partial_j)$, T_{ijk} é simétrico nos índices j e k . □

É importante salientar que uma estrutura estatística conduz a uma estrutura dualística, como expressa o seguinte teorema.

Teorema 4. (AY *et al.*, 2017) Uma métrica riemanniana g e um tensor 3-simétrico em uma variedade diferenciável M induz uma estrutura dualística livre de torção em M .

Demonstração. Seja $\nabla^{(0)}$ a conexão de Levi-Civita de g . Defina, para todo $A, B, C \in \mathfrak{X}(M)$, a conexão ∇ por

$$\nabla_A B := \nabla_A^{(0)} B - \frac{1}{2} \mathcal{H}(A, B),$$

onde \mathcal{H} é definida por

$$g(\mathcal{H}(A, B), C) = T(A, B, C).$$

Desse modo, observamos que

$$g(\nabla_A B, C) = g\left(\nabla_A^{(0)} B, C\right) - \frac{1}{2} T(A, B, C),$$

de maneira que $\nabla_A B$ é linear sobre funções suaves em A e sobre escalares reais em B . Além disso, para toda função suave em M ,

$$g(\mathcal{H}(A, fB), C) = T(A, fB, C) = fT(A, B, C) = g(f\mathcal{H}(A, B), C)$$

e, conseqüentemente,

$$\begin{aligned} \nabla_A(fB) &= \nabla_A^{(0)}(fB) - \frac{1}{2}\mathcal{H}(A, fB) \\ &= \frac{\partial f}{\partial A}B + f\nabla_A^{(0)}B - \frac{f}{2}\mathcal{H}(A, B), \\ &= \frac{\partial f}{\partial A}B + f\nabla_A B, \end{aligned}$$

ou seja, ∇ segue a regra do produto para conexões. Assim, ∇ é uma conexão linear em M . Perceba que, como

$$g(\mathcal{H}(A, B), C) = T(A, B, C) = T(B, A, C) = g(\mathcal{H}(B, A), C)$$

para todo $A, B, C \in \mathfrak{X}(M)$, \mathcal{H} é simétrico. Como consequência, temos que o tensor de torção τ de ∇ é

$$\begin{aligned} \tau(A, B) &= \nabla_A B - \nabla_B A - [A, B] - \frac{1}{2}(\mathcal{H}(A, B) - \mathcal{H}(B, A)) \\ &= \nabla_A^{(0)}B - \nabla_B^{(0)}A - [A, B] \\ &= 0 \end{aligned}$$

para todo $A, B \in \mathfrak{X}(M)$. Defina, agora, a conexão ∇^* em M , também para todo $A, B \in \mathfrak{X}(M)$, dada por

$$\nabla_A^* B = \nabla_A^{(0)}B + \frac{1}{2}\mathcal{H}(A, B).$$

Pelos mesmos argumentos utilizados para ∇ , ∇^* é uma conexão linear livre de torção em M . Além disso,

$$\begin{aligned} Ag(B, C) &= g\left(\nabla_A^{(0)}B, C\right) + g\left(B, \nabla_A^{(0)}C\right) \\ &= g\left(\nabla_A^{(0)}B, C\right) + g\left(B, \nabla_A^{(0)}C\right) - \frac{1}{2}T(A, B, C) + \frac{1}{2}T(A, C, B) \\ &= g(\nabla_A B, C) + g(B, \nabla_A^* C) \end{aligned}$$

para todo $A, B, C \in \mathfrak{X}(M)$, ou seja, ∇^* é dual a ∇ em relação a g . \square

Desse modo, observamos que existe uma relação bidirecional entre uma estrutura dualística e uma estrutura estatística. Há autores, como os de (AY *et al.*, 2017), que classificam essa última como “mais compacta”⁷ que a primeira, permitindo a definição de uma família de conexões livres de torção e indexadas por $\alpha \in [-1, 1]$, dadas por

$$\nabla_A^{(\alpha)} B := \nabla_A^{(0)} B - \frac{\alpha}{2}\mathcal{H}(A, B), \quad (4.6)$$

⁷ Por “mais compacta”, entende-se que (g, T) é mais geral que (g, ∇, ∇^*) , induzindo não apenas uma, mas infinitamente muitas estruturas dualísticas em M .

para todo $A, B \in \mathfrak{X}(M)$, sendo essa família a das já conhecidas α -conexões. Pelos mesmos argumentos desenvolvidos na demonstração de 4, $\nabla^{(\alpha)}$ é, de fato, uma conexão livre de torção em M cuja conexão dual é dada por

$$\nabla_A^{(-\alpha)} B := \nabla_A^{(0)} B + \frac{\alpha}{2} \mathcal{H}(A, B).$$

Outrossim, é claro que

$$\nabla^{(0)} = \frac{1}{2} \left(\nabla^{(0)} + \frac{\alpha}{2} \mathcal{H} + \nabla^{(0)} - \frac{\alpha}{2} \mathcal{H} \right) = \frac{\nabla^{(\alpha)} + \nabla^{(-\alpha)}}{2},$$

ou seja, a conexão de Levi-Civita de g é a média entre α -conexões duais.

É relevante destacar que em algumas literaturas, como em (UOHASHI, 2017) e (NIELSEN, 2020), a definição para as α -conexões é apresentada com $\alpha \in \mathbb{R}$. Aqui, optamos por empregar aquela proposta em (AY *et al.*, 2017). A escolha por restringir o domínio do parâmetro α a $[-1, 1]$ não é arbitrária. As conexões $\nabla^{(-1)}$ e $\nabla^{(1)}$ são destacadas, sendo as já conhecidas conexões mistura e exponencial, respectivamente, apresentadas no contexto de geometria da informação finita. Assim, $\alpha \in [-1, 1]$ já contempla o comportamento geral da geometria advinda das α -conexões.

4.2 As famílias exponenciais e mistura - um motivação para a geometria da informação plana

Como foi possível observar em 2, algumas famílias de distribuições de probabilidade destacam-se, especialmente sob a óptica da geometria riemanniana. É fato que parte dessa geometria preocupa-se com o estudo das conexões planas. Assim, nesta seção, apresentaremos duas classes de famílias de distribuições que possuem um conteúdo geométrico particular, capaz de estimular o estudo de uma geometria da informação plana⁸. Começemos, então, apresentando as famílias exponenciais.

Definição 23 (Família exponencial). (AY *et al.*, 2017) Uma família exponencial em um conjunto \mathcal{X} é⁹ uma família de distribuições de probabilidade da forma

$$p(x; \vartheta) = \exp \left(\gamma(x) + \sum_{i \in J} f_i(x) \vartheta^i - \psi(\vartheta) \right) \mu(x),$$

onde $\vartheta = (\vartheta^1, \dots, \vartheta^n)$ é um parâmetro n -dimensional de um espaço de parâmetros Ξ , $\gamma(x)$ e $f_1(x), \dots, f_n(x)$ são funções em \mathcal{X} , $\mu(x)$ é uma medida em \mathcal{X} e ψ é função dos parâmetros

⁸ Por mais que a construção do texto deixe claro que, com “geometria da informação plana”, estamos nos referindo a uma geometria da informação advinda de conexões planas, esta seção e as subseqüentes deixarão mais evidente o que isso significa.

⁹ É importante observar que \mathcal{X} é tomado munido de uma σ -álgebra.

dada por

$$\psi(\vartheta) = \ln \int_{\mathcal{X}} \exp \left(\gamma(x) + \sum_{i \in J} f_i(x) \vartheta^i \right) d\mu(x) \quad (4.7)$$

para garantir normalização de p em \mathcal{X} .

Perceba que o espaço dos parâmetros ϑ , neste caso, corresponde ao conjunto dos ϑ tais que a integral no argumento do logaritmo de 4.7 é finita. O principal exemplo de uma família exponencial é o das distribuições gaussianas $\mathcal{N}(\mu, \sigma)$ com medida de Lebesgue. Como

$$\mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} = \exp \left(\frac{x\mu}{\sigma^2} - \frac{x^2}{2\sigma^2} - \frac{\mu^2}{2\sigma^2} - \ln \sqrt{2\pi\sigma} \right),$$

observamos que

$$\gamma(x) = 0, \quad f_1(x) = x, \quad f_2(x) = x^2, \quad \vartheta^1 = \frac{\mu}{\sigma^2}, \quad \vartheta^2 = -\frac{1}{2\sigma^2}, \quad \psi(\vartheta) = \frac{\mu^2}{2\sigma^2} + \ln \sqrt{2\pi\sigma}.$$

Um segundo exemplo são as distribuições de Poisson, tomadas sobre os inteiros não negativos $\mathbb{Z}_{\geq 0}$ com a medida de contagem. Nesse caso,

$$p(\xi; x) = e^{-\xi} \frac{\xi^x}{x!} = \exp(-\ln x! + x \ln \xi - \xi),$$

de forma que fica claro que

$$\gamma(x) = -\ln x!, \quad f_1(x) = x, \quad \vartheta^1 = \ln \xi, \quad \psi(\vartheta) = \xi.$$

Ademais, lembrando que $\nabla^{(1)}$ é chamada de conexão exponencial, vamos a analisar no contexto das famílias exponenciais. Para tal, definamos o que é uma conexão plana.

Definição 24 (Conexão plana). (NIELSEN, 2022) Seja M uma variedade diferenciável munida de uma conexão linear ∇ . Dizemos que ∇ é plana se ela é livre de torção e se existe um sistema de coordenadas local tal que os símbolos de Christoffel dela são nulos quando expressos nesse sistema, que recebe o nome de um sistema de coordenadas afim para ∇ .

Definamos, também, o conceito de paralelismo de campos de vetores.

Definição 25 (Campos paralelos de vetores). (AY *et al.*, 2017) Seja M uma variedade diferenciável munida de uma conexão linear ∇ . Dada uma vizinhança $U \subset M$ de $p \in M$, um campo de vetores em U é dito paralelo se, para qualquer outro campo de vetores W em U , tivermos que $\nabla_W V = 0$.

Note, então, que uma consequência direta de 24 envolve paralelismo de referenciais locais. De fato, dado um sistema de coordenadas afim $\vartheta = (\vartheta^1, \dots, \vartheta^n)$ para uma conexão ∇ em uma vizinhança $U \subset M$ de $p \in M$, devido à nulidade dos símbolos de Christoffel de ∇ nesse sistema, os campos canônicos $\frac{\partial}{\partial \vartheta^1}, \frac{\partial}{\partial \vartheta^2}, \dots, \frac{\partial}{\partial \vartheta^n}$ são campos paralelos em U .

Colocado isso e dotados da definição 24, sigamos estudando $\nabla^{(1)}$.

Proposição 9. (AY *et al.*, 2017) A conexão $\nabla^{(1)}$ em uma família exponencial é plana.

Demonstração. Note que, para todo $i \in J$, vale a sequência de igualdades

$$\begin{aligned} \partial_i \ln p_{\vartheta}(x) &= \frac{\partial_i p_{\vartheta}(x)}{p_{\vartheta}(x)} \\ &= f_i(x) - \frac{\int_{\mathcal{X}} \exp(\gamma(x) + \sum_{j \in J} f_j(x) \vartheta^j) f_i(x) \, d\mu(x)}{\int_{\mathcal{X}} \exp(\gamma(x) + \sum_{j \in J} f_j(x) \vartheta^j) \, d\mu(x)} \\ &= f_i(x) - \frac{\int_{\mathcal{X}} \exp(\gamma(x) + \sum_{j \in J} f_j(x) \vartheta^j) f_i(x) \, d\mu(x)}{e^{\psi(x)}} \\ &= f_i(x) - \mathbf{E}_{\vartheta} [f_i], \end{aligned}$$

de maneira que

$$\mathbf{E}_{\vartheta} [\partial_i \ln p_{\vartheta}(x)] = \mathbf{E}_{\vartheta} [f_i(x) - \mathbf{E}_{\vartheta} (f_i)] = \mathbf{E}_{\vartheta} [f_i] - \mathbf{E}_{\vartheta} [f_i] = 0. \quad (4.8)$$

Além disso, perceba que, para todo $i, j \in J$,

$$\begin{aligned} \partial_i \partial_j \ln p_{\vartheta}(x) &= \partial_i (f_j(x) - \mathbf{E}_{\vartheta} [f_j]) \\ &= -\partial_i \left[\frac{\int_{\mathcal{X}} \exp(\gamma(x) + \sum_{m \in J} f_m \vartheta^m) f_j \, d\mu(x)}{\int_{\mathcal{X}} \exp(\gamma(x) + \sum_{m \in J} f_m \vartheta^m) \, d\mu(x)} \right] \\ &= -\partial_i \left[\partial_j \left(\ln \int_{\mathcal{X}} \exp \left(\gamma(x) + \sum_{m \in J} f_m \vartheta^m \right) \, d\mu(x) \right) \right], \end{aligned}$$

ou seja,

$$\partial_i \partial_j \ln p_{\vartheta}(x) = -\partial_i \partial_j \psi(\vartheta). \quad (4.9)$$

Substituindo tanto essa expressão quanto 4.8 em 2.8, obtemos

$$\begin{aligned} \Gamma_{ijk}^{(1)} &= \mathbf{E}_{\vartheta} (\partial_i \partial_j \ln p_{\vartheta}(x) \partial_k \ln p_{\vartheta}(x)) \\ &= -\partial_i \partial_j \psi(\vartheta) \mathbf{E}_{\vartheta} [\partial_k \ln p_{\vartheta}(x)] \\ &= 0, \end{aligned}$$

para todo $i, j, k \in J$. Assim, como $\nabla^{(1)}$ é livre de torção, ela é plana em famílias exponenciais e $(\vartheta^1, \vartheta^2, \dots, \vartheta^n) \in \mathbb{R}^n \mapsto p_{\vartheta}$ é um sistema de coordenadas afim para essa conexão. \square

Além disso, note que, usando 4.9 e 2.5, podemos escrever a métrica de Fisher nesse sistema afim como

$$g_{ij}(\vartheta) = \int_{\mathcal{X}} \partial_i \partial_j \psi(\vartheta) \, d\mu(x) = \partial_i \partial_j \psi(\vartheta). \quad (4.10)$$

Outrossim, consideremos agora uma segunda família de distribuições, chamadas de distribuições mistura.

Definição 26 (Família mistura). (AY *et al.*, 2017) Uma família mistura em um conjunto \mathcal{X} é uma família de distribuições de probabilidade da forma

$$p(x; \eta) = c(x) + \sum_{i \in J} g^i(x) \eta_i,$$

onde $\eta = (\eta_1, \eta_2, \dots, \eta_n)$ é um parâmetro n -dimensional de um espaço de parâmetros Ξ , $c(x)$ e $g^1(x), \dots, g^n(x)$ são funções em \mathcal{X} tais que¹⁰

$$\int_{\mathcal{X}} c(x) \, dx = 1, \quad \int_{\mathcal{X}} g^i(x) \, dx = 0$$

para todo $i \in J$.

A nomenclatura para essa família de distribuições advém delas serem definidas como combinação ou, em outras palavras, uma mistura de distribuições de probabilidade g^i em \mathcal{X} não necessariamente parametrizadas. Por isso, algumas vezes c é desconsiderada na definição acima, de maneira que uma família mistura pode ser escrita como soma convexa das g^i . Agora, seguindo tal qual foi feito para as famílias exponenciais, vamos estudar a conexão mistura $\nabla^{(-1)}$ neste caso.

Proposição 10. (AY *et al.*, 2017) A conexão $\nabla^{(-1)}$ em uma família mistura é plana.

Demonstração. Primeiramente, para todo $i \in J$, temos que¹¹

$$\partial^i \ln p_\eta(x) = \frac{\partial^i p_\eta(x)}{p_\eta(x)} = \frac{g^i(x)}{p_\eta(x)},$$

de forma que

$$\partial^i \partial^j \ln p_\eta(x) = \partial^i \left(\frac{g^j(x)}{p_\eta(x)} \right) = -\frac{g^i(x) g^j(x)}{p_\eta(x)^2}. \quad (4.11)$$

Assim, para todo $i, j \in J$,

$$\partial^i \partial^j \ln p_\eta(x) + (\partial^i \ln p_\eta(x)) (\partial^j \ln p_\eta(x)) = -\frac{g^j(x) g^i(x)}{p_\eta(x)^2} + \left(\frac{g^i(x)}{p_\eta(x)} \right) \left(\frac{g^j(x)}{p_\eta(x)} \right) = 0,$$

que, substituindo em 2.8, fornece $\Gamma_{ijk}^{(-1)} = 0$ para todo $i, j, k \in J$. Logo, como $\nabla^{(-1)}$ é livre de torção, ela é plana em famílias mistura e $(\eta_1, \eta_2, \dots, \eta_n) \in \mathbb{R}^n \mapsto p_\eta$ é um sistema de coordenadas afim para essa conexão. \square

¹⁰ Essa hipótese sobre c e g^i , $i \in J$, pode ser retirada somando o termo

$$v(\eta) := 1 - \int_{\mathcal{X}} \left(c(x) + \sum_{i \in J} g^i(x) \eta_i \right) dx$$

na expressão para $p(x; \eta)$ (AY *et al.*, 2017). Além disso, c costuma ser chamada de função cumulante ou, em física estatística, de logaritmo da função de partição (NIELSEN, 2022).

¹¹ Analogamente à notação ∂_i que vinha sendo usada para $\frac{\partial}{\partial \xi^i}$ e que foi estendida para representar $\frac{\partial}{\partial \eta^i}$, ∂^i denota $\frac{\partial}{\partial \eta^i}$.

Posto isso, usemos 4.11 para calcular a métrica de Fisher nesse sistema. Por 2.5,

$$\begin{aligned} g_{ij}(\eta) &= \int_{\mathcal{X}} \frac{g^i(x)g^j(x)}{p_\eta(x)} dx \\ &= \int_{\mathcal{X}} \frac{\partial^i p_\eta(x) \partial^j p_\eta(x)}{p_\eta(x)} dx \\ &= \int_{\mathcal{X}} \partial^i \partial^j p_\eta(x) \ln p_\eta(x) dx + \int_{\mathcal{X}} \frac{\partial^j p_\eta(x) \partial^i p_\eta(x)}{p_\eta(x)} dx, \end{aligned}$$

isto é,

$$g_{ij}(\eta) = -\partial^i \partial^j h(p_\eta), \quad (4.12)$$

onde h corresponde à entropia de Shannon de p_η , já citada em 2.2.2 e dada por

$$h(p_\eta) = - \int_{\mathcal{X}} p_\eta(x) \ln p_\eta(x) dx.$$

Então, observando tanto 4.11 quanto 4.12, é possível notar que não só os modelos exponenciais, mas também os mistura possuem métricas que são obtidas a partir de hessianas de funções: $\text{Hess}(\psi(\vartheta))$ e $\text{Hess}(h(p_\eta))$, respectivamente, onde a notação $\text{Hess}(f)$ corresponde à hessiana da função f . Isso expressa o fato de que essas famílias podem ser modeladas por variedades hessianas (NIELSEN, 2022).

Definição 27 (Variedade hessiana). (SHIMA; YAGI, 1997) Seja M uma variedade diferenciável munida de uma conexão linear plana ∇ . Uma métrica g em M é dita uma métrica hessiana se ela é localmente expressa como a hessiana de uma função local suave. Chamamos uma tripla ordenada (M, ∇, g) com g sendo uma métrica hessiana de uma variedade hessiana.

Juntando isso ao fato de $\nabla^{(-1)}$ e $\nabla^{(1)}$ serem conexões duais que são planas em cada respectivo modelo apresentado, é legítimo destacar o contexto em que ambas são planas em uma variedade hessiana.

4.3 A dualidade em geometria da informação e as transformadas de Legendre

4.3.1 Introdução a variedades dualmente planas

Na seção anterior, recorreremos às famílias exponenciais e mistura para fomentarmos o estudo da dualidade em um contexto de conexões planas, dando destaque aquele referente às variedades hessianas. Além de fornecerem mais justificativas para a nomenclatura adotada para $\nabla^{(-1)}$ e $\nabla^{(1)}$, os exemplos apresentados em 4.2 levam-nos a estudar variedades munidas de uma métrica que seja hessiana. Dito isso, podemos começar esse estudo definindo o que é uma estrutura dualmente plana em uma variedade diferenciável.

Definição 28 (Estrutura dualmente plana). (AY *et al.*, 2017) Seja (M, g, ∇, ∇^*) variedade dualística. A tripla ordenada (g, ∇, ∇^*) é chamada uma estrutura dualmente plana para M se é uma estrutura dualística para M tal que tanto ∇ quanto ∇^* são planas.

Primeiramente, podemos nos perguntar quais são as subvariedades de M que, munidas da estrutura dualística induzida de (M, ∇, ∇^*) , também são dualmente planas. Esse é um questionamento complexo, que pode ser respondido parcialmente a partir da próxima proposição, que garante que subvariedades autoparalelas de M expressam tal comportamento. Por conseguinte, definamos o que são elas.

Definição 29 (Subvariedade autoparalela). (AY *et al.*, 2017; OHARA, 2019) Seja S uma subvariedade¹² de M . Sendo M munida de uma métrica riemanniana g e de uma conexão linear ∇ , S é dita autoparalela¹³ em M em relação a ∇ se, para todo $A, B \in \mathfrak{X}(S)$, tivermos que $\nabla_A B \in \mathfrak{X}(S)$.

Isso apresentado, sigamos para a proposição que vincula autoparalelismo¹⁴ com dualidade plana no contexto de subvariedades.

Proposição 11. Seja (g, ∇, ∇^*) uma estrutura dualmente plana na variedade M e seja, também, S uma subvariedade de M . Se S é autoparalela para ∇ ou ∇^* , então ela é dualmente plana com respeito à métrica e conexões induzidas de (g, ∇, ∇^*) .

Demonstração. Sem perda de generalidade, vamos supor que S é autoparalela em relação a ∇ , uma vez que todos os cálculos podem ser igualmente feitos no caso de autoparalelismo com respeito a ∇^* . Assim, a conexão induzida por ∇ em S coincide com ela própria restrita a campos de vetores tangentes a S e, por consequência, é plana. Além disso, denotando por g^S a métrica em S herdada de (M, g) , pela definição de conexões duais, observamos que a conexão $\nabla^{S,*}$ induzida por ∇^* em S deve satisfazer

$$g^S(\nabla_C A, B) + g^S(A, \nabla_C^{S,*} B) = Cg^S(A, B) = Cg(A, B) = g(\nabla_C A, B) + g(A, \nabla_C^* B) \quad (4.13)$$

para todo $A, B, C \in \mathfrak{X}(S)$. Como, por hipótese, ∇^* é livre de torção, então observamos por 4.13 que $\nabla^{S,*}$ também é. Além disso, uma vez que ∇ possui tensor de curvatura nulo, afirmamos que $\nabla^{S,*}$ também tem. Esse é um resultado que será provado posteriormente¹⁵, mas corresponde a 12. Logo, $\nabla^{S,*}$ é conexão plana e S é dualmente plana com respeito a (g^S, ∇, ∇^*) . \square

¹² Para (AY *et al.*, 2017), subvariedades correspondem, na verdade, a subvariedades suaves ou também chamadas de regulares. Essa será a convenção que usaremos ao estudar subvariedades.

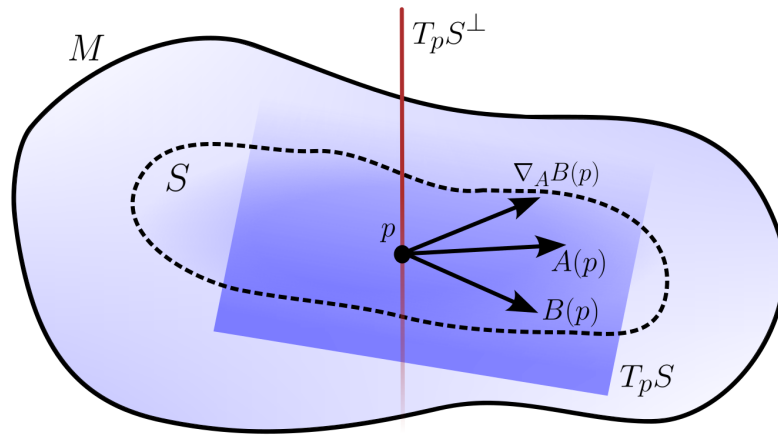
¹³ As vezes se utiliza a denominação ∇ -autoparalela a M para identificar a dependência com a conexão.

¹⁴ Aqui, usamos autoparalelismo como o atributo que define variedades autoparalelas em um contexto em que se tem, subentendido, a variedade M e a conexão linear.

¹⁵ A escolha de apresentar um resultado que necessita de um segundo que ainda será demonstrado foi uma decisão tomada visando manter uma sequência lógica no texto que se estende para além da organização desses resultados. Caso essa decisão fosse diferente, 11 ficaria descontextualizada no corpo deste trabalho.

Assim, é perceptível que o autoparalelismo é um atributo importante de subvariedades dentro do contexto de geometria da informação plana, uma vez que preserva estruturas dualmente planas. Convém ressaltar que tal consciência conduziu à rogativa de hipóteses ainda mais fortes com respeito ao autoparalelismo, solicitando que uma subvariedade fosse autoparalela em relação não a uma única conexão, mas a duas conexões que compõem uma estrutura dualística para a variedade ambiente. A essas subvariedades, como esperado, deu-se o nome de variedades duplamente autoparalelas (OHARA, 2019).

Figura 11 – Representação gráfica de uma subvariedade autoparalela S de M . Denotando por $T_p S^\perp$ o complemento ortogonal de $T_p S$ em $T_p M$, como $\nabla_{AB} \in \mathfrak{X}(S)$ para todo $A, B \in S$, a projeção de $\nabla_{AB}(p)$ em $T_p S^\perp$ é nula para todo $p \in S$, o que motiva a nomenclatura para S



Fonte: Elaborada pelo autor.

Outrossim, considere, agora, M variedade diferenciável munida de uma estrutura dualmente plana (g, ∇, ∇^*) . Como ∇ é plana, existem coordenadas afim $\vartheta^1, \vartheta^2, \dots, \vartheta^n$ para ela. Consequentemente, $\frac{\partial}{\partial \vartheta^1}, \frac{\partial}{\partial \vartheta^2}, \dots, \frac{\partial}{\partial \vartheta^n}$ são campos paralelos em M . A partir deles, defina n campos de vetores ∂^j também em M dados através da relação

$$g(\partial_i, \partial^j) = \delta_i^j, \quad i, j \in J, \quad (4.14)$$

onde δ_i^j representa a delta de Kronecker, de forma que ∂^j ficam bem definidos. Por meio da dualidade de ∇ e ∇^* unida à expressão 4.14, sabemos que, para todo $V \in \mathfrak{X}(M)$ e para todo $i, j \in J$,

$$g(\nabla_V \partial_i, \partial^j) + g(\partial_i, \nabla_V^* \partial^j) = V g(\partial_i, \partial^j) = 0.$$

Consequentemente, para todo $j \in J$, $\nabla_V^* \partial^j = 0$, ou seja, ∂^j são campos paralelos para ∇^* . Por conseguinte, existe sistema de coordenadas afim $\eta_1, \eta_2, \dots, \eta_n$ para ∇^* tal que $\partial^j = \frac{\partial}{\partial \eta_j}$ para todo $j \in J$. Desse modo, fica claro que a transição das coordenadas ϑ para as η e vice-versa é fornecida através das relações

$$\partial^j = \sum_{i \in J} (\partial^j \vartheta^i) \partial_i = \sum_{i \in J} \left(\frac{\partial \vartheta^i}{\partial \eta_j} \right) \frac{\partial}{\partial \vartheta^i},$$

$$\partial_i = \sum_{j \in J} (\partial_i \eta_j) \partial^j = \sum_{j \in J} \left(\frac{\partial \eta_j}{\partial \vartheta^i} \right) \frac{\partial}{\partial \eta_j}.$$

Dessa maneira, escrevendo g nas coordenadas ϑ e η , respectivamente, obtemos

$$g_{ij} := g(\partial_i, \partial_j) = g \left(\sum_{k \in J} \left(\frac{\partial \eta_k}{\partial \vartheta^i} \right) \frac{\partial}{\partial \eta_k}, \frac{\partial}{\partial \vartheta^j} \right) = \frac{\partial \eta_j}{\partial \vartheta^i},$$

$$g^{ij} := g(\partial^i, \partial^j) = g \left(\sum_{k \in J} \left(\frac{\partial \vartheta^k}{\partial \eta_i} \right) \frac{\partial}{\partial \vartheta^k}, \frac{\partial}{\partial \eta_j} \right) = \frac{\partial \vartheta^j}{\partial \eta_i}.$$

Sabemos que, dada uma função estritamente convexa suave¹⁶ ψ de ϑ em um subconjunto aberto de M , a hessiana $\text{Hess}(\psi(\vartheta))$ de ψ é uma matriz positiva definida dependente de ϑ e que define uma métrica riemanniana. De fato, tomando o termo quadrático da expansão em série de Taylor de ψ em torno de ϑ , dado por $\frac{1}{2} d\vartheta^T \text{Hess}(\psi(\vartheta)) d\vartheta$, estabelecemos a matriz que define uma métrica g no referencial formado pelos campos coordenados ∂_i como $\text{Hess}(\psi(\vartheta))$ (AMARI, 2009). Por outro lado, retornando ao cenário de variedades dualmente planas, vamos mostrar que a recíproca é verdadeira, ou seja, que dados sistemas de coordenadas afim para ∇ e ∇^* em (M, g, ∇, ∇^*) variedade dualística plana, existe φ função convexa tal que foi apresentado.

Teorema 5. Existem funções potenciais estritamente convexas $\varphi : M \rightarrow \mathbb{R}$ e $\psi : M \rightarrow \mathbb{R}$ satisfazendo

$$\eta_i = \partial_i \psi(\vartheta), \quad \vartheta^i = \partial^i \varphi(\eta),$$

assim como

$$g_{ij} = \partial_i \partial_j \psi, \quad g^{ij} = \partial^i \partial^j \varphi.$$

Demonstração. Observe que a equação diferencial $\eta_i = \partial_i \psi$ possui solução local se, e somente se,

$$\partial_i \eta_j = \partial_j \eta_i. \quad (4.15)$$

Desse modo, existe ψ que satisfaz a equação diferencial parcial em questão para todo $i \in J$. Como consequência, escrevendo g novamente no sistema de coordenadas ϑ , obtemos que

$$g_{ij} = \partial_i \eta_j = \partial_i \partial_j \psi$$

para todo $i, j \in J$ e, desse modo, verificamos que ψ é estritamente convexa¹⁷. Defina, agora,

$$\varphi := \sum_{i \in J} \vartheta^i \eta_i - \psi. \quad (4.16)$$

¹⁶ A partir daqui, iremos suprimir o adjetivo suave, uma vez que toda função convexa, neste trabalho, terá tal atributo.

¹⁷ Note que poderíamos utilizar essa mesma argumentação para mostrar a existência de φ , contudo, utilizaremos a dualidade entre os sistemas de coordenadas. A justificativa para essa escolha é a utilidade que isso terá para nós nos passos seguintes deste trabalho.

Então,

$$\begin{aligned}\partial^i \varphi &= \vartheta^i + \sum_{j \in J} \frac{\partial \vartheta^j}{\partial \eta_i} \eta_j - \sum_{j \in J} \frac{\partial \vartheta^j}{\partial \eta_i} \frac{\partial \psi}{\partial \eta_j} \\ &= \vartheta^i + \sum_{j \in J} \frac{\partial \vartheta^j}{\partial \eta_i} \frac{\partial \psi}{\partial \eta_j} - \sum_{j \in J} \frac{\partial \vartheta^j}{\partial \eta_i} \frac{\partial \psi}{\partial \eta_j} \\ &= \vartheta^i.\end{aligned}$$

Assim, φ é solução para a equação diferencial parcial $\vartheta^i = \partial^i \varphi$. Logo, escrevendo g^{-1} no sistema de coordenadas η , obtemos que

$$g^{ij} = \partial^i \vartheta^j = \partial^i \partial^j \varphi$$

para todo $i, j \in J$ e, conseqüentemente, φ é estritamente convexa. \square

Observando a expressão 4.16 sob a perspectiva da análise convexa, fica evidente que φ e ψ relacionam-se através de uma transformada de Legendre, que exerce função destacada no estudo na geometria da informação plana (NIELSEN, 2010).

4.3.2 Uma interpretação intuitiva sobre a transformada de Legendre

A fim de compreender a transformada de Legendre sob uma perspectiva geométrica e, em certo grau, intuitiva, estudemos o caso de funções convexas definidas em \mathbb{R} , seguindo o que foi apresentado em (NIELSEN, 2010).

Seja $F : U \rightarrow \mathbb{R}$ uma função estritamente convexa definida em um aberto $U \subset \mathbb{R}^n$. Considere, também, os conjuntos

$$\mathcal{G}(F) := \{(x, F(x)) \mid x \in U\},$$

$$\mathcal{O}(F) := \{(x, z) \mid x \in U, z \geq F(x)\},$$

chamados, respectivamente, de gráfico e de epígrafe de F . Examinando o bordo $\partial \mathcal{O}$ de \mathcal{O} , podemos estudar como F está codificada nele. Com tal propósito, para cada $P \in \partial \mathcal{O}$, podemos tomar o sistema de coordenadas natural X dado pela projeção ortogonal, isto é, $X(P) = (x(P), F(x(P)))$. Em P , temos o hiperplano tangente H_P , de equação conhecida e dada por

$$z = (x - x(P))^T \text{grad}_{x(P)}(F) + F(x(P)),$$

onde $\text{grad}_{x(P)}(F)$ fornece quão inclinado H_P está em relação ao plano $(x, 0)$. Por um lado, é trivial que cada P define unicamente $\text{grad}_{x(P)}(F)$. Por outro lado, como F é convexa, cada entrada de $\text{grad}(F)$ é uma função monotonicamente crescente e, por consequência, cada ponto $P \in \partial \mathcal{O}$ é determinado de forma única por $\text{grad}_{x(P)}(F)$, de maneira que esse gradiente define um sistema de coordenadas Y dado por $Y = \text{grad}_x(F)$. Esses dois sistemas formam o que é chamado de um sistema de coordenadas dual.

Isso posto, podemos nos questionar como escrever os pontos $P \in \partial \mathcal{O}$ no sistema Y . Primeiramente, note que cada inclinação y determina uma família de hiperplanos H_Q dada por

$$z = (x - x(Q))^T y + F(x(Q)),$$

com $Q \in \partial \mathcal{O}$. Além disso, perceba que cada H_P intercepta o eixo z em um único ponto, de tal modo que H_P é o único plano que minimiza $z = (0 - x(P))^T \text{grad}_{x(P)}(F) + F(x(P))$. Logo, temos que P pode ser obtido por

$$P = \arg \min_{Q \in \partial \mathcal{O}} \{-x(Q)^T y + F(x(Q))\} = \arg \max_{Q \in \partial \mathcal{O}} \{x(Q)^T y - F(x(Q))\}.$$

Grosso modo, o que $\partial \mathcal{O}$ expressa é o formato da função F , sendo determinado pelo domínio e conjunto imagem desta última. Assim, podemos definir

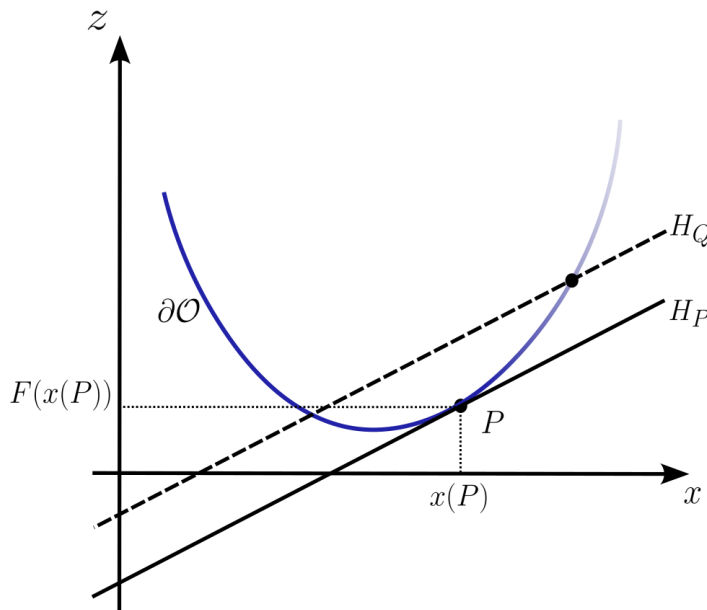
$$G(y) := \max_{x \in U} \{x^T y - F(x)\},$$

parametrizado por $y = \text{grad}_x(F) \in V$, como uma outra maneira de codificar o formato de F , onde

$$V := \{\text{grad}_x(F) \mid x \in U\}.$$

Como G é igualmente estritamente convexa, seu formato pode ser codificado de maneira semelhante por F .

Figura 12 – Representação gráfica de P como argumento de minimização da intersecção de H_Q com o eixo z , para Q ponto do gráfico da função estritamente convexa F e H_Q hiperplano de inclinação $\text{grad}_{x(P)}(F)$ passando por Q



Fonte: Adaptada de Nielsen (2010, p 2).

Então, apresentado isso e transportando tal noção para o contexto de 5, observamos que φ e ψ configuram um sistema de coordenadas dual, relacionando-se pelas transformadas de Legendre

$$\varphi(\eta) = \max_{\vartheta} \left(\sum_{i \in J} \vartheta^i \eta_i - \psi(\vartheta) \right), \quad \psi(\vartheta) = \max_{\eta} \left(\sum_{i \in J} \vartheta^i \eta_i - \varphi(\eta) \right),$$

que são válidas mesmo localmente.

4.4 A geometria da dualidade em geometria da informação plana

Como foi citado na seção anterior, uma função convexa¹⁸ $\psi(\vartheta)$ induz uma métrica riemanniana em M . Para finalizar este trabalho, vamos explorar, aqui, a geometria que surge quando partimos de ψ . Primeiramente, como foi citado em 4.3, definamos a métrica g em M a partir da hessiana de ψ . Assim, g fica determinada na base formada pelos campos $\partial_i = \frac{\partial}{\partial \vartheta^i}$, com $i \in J$, pela relação $g_{ij} = \partial_i \partial_j \psi$. De maneira análoga, definimos o tensor 3-simétrico T em M através da expressão $T_{ijk} = \partial_i \partial_j \partial_k \psi$. Note que, pelo teorema de Schwarz para simetria das derivadas segundas, T é, de fato, simétrico em suas entradas. Além disso, pela linearidade dos campos ∂_i agindo como derivações, T é trilinear, sendo, efetivamente, um campo tensorial 3-covariante em M .

Ademais, usando 4.6, sabemos que

$$g \left(\nabla_i^{(\alpha)} \partial_j, \partial_k \right) = g \left(\nabla_i^{(0)} \partial_j, \partial_k \right) - \frac{\alpha}{2} g \left(\mathcal{H}(\partial_i, \partial_j), \partial_k \right) = g \left(\nabla_i^{(0)} \partial_j, \partial_k \right) - \frac{\alpha}{2} T_{ijk},$$

de maneira que os símbolos de Christoffel (a menos de abaixamento de um índice) que definem as α -conexões em M são dados por

$$\Gamma_{ijk}^{(\alpha)} = \Gamma_{ijk}^{(0)} - \frac{\alpha}{2} \partial_i \partial_j \partial_k \psi, \quad (4.17)$$

para $\alpha \in [-1, 1]$. Lembrando que

$$\Gamma_{ijk}^{(0)} = \frac{1}{2} \{ \partial_i g_{jk} + \partial_j g_{ki} - \partial_k g_{ij} \},$$

que, neste caso, resume-se a

$$\Gamma_{ijk}^{(0)} = \frac{1}{2} \partial_i \partial_j \partial_k \psi,$$

podemos substituir essa expressão em 4.17, obtendo, assim,

$$\Gamma_{ijk}^{(\alpha)} = \frac{(1-\alpha)}{2} \partial_i \partial_j \partial_k \psi. \quad (4.18)$$

¹⁸ Durante alguns momentos desta seção, quando não houver confusão, iremos escrever apenas ψ , a fim de manter a notação mais sucinta.

Dessa maneira, fica claro que $\nabla^{(1)}$, sendo livre de torção, é uma conexão plana, expressando o fato de que ϑ configura-se como um sistema de coordenadas afim para ela. Também é possível perceber que a conexão de Levi-Civita, como esperado, escreve-se como a média entre as α -conexões com α opostos, isto é,

$$\Gamma_{ijk}^{(0)} = \frac{\Gamma_{ijk}^{(0)} - \frac{\alpha}{2} \partial_i \partial_j \partial_k \psi + \Gamma_{ijk}^{(0)} + \frac{\alpha}{2} \partial_i \partial_j \partial_k \psi}{2} = \frac{\Gamma_{ijk}^{(\alpha)} + \Gamma_{ijk}^{(-\alpha)}}{2}.$$

Outrossim, calculemos agora o endomorfismo de curvatura de Riemann¹⁹. Usualmente, ele é definido a partir da conexão de Levi-Civita, contudo, como ela é contemplada pelas α -conexões, calculemos ele a partir delas. Realizando levantamento de índice, obtemos que os símbolos de Christoffel de $\nabla^{(\alpha)}$ são

$$\Gamma_{ij}^{(\alpha)k} = \sum_{m \in J} \Gamma_{ijm}^{(\alpha)} g^{mk} = \frac{(1-\alpha)}{2} \sum_{m \in J} \partial_i \partial_j \partial_m \psi g^{mk}.$$

Consequentemente²⁰,

$$\begin{aligned} R_{lij}^{(\alpha)k} &= \sum_m \left(\partial_i \Gamma_{jl}^{(\alpha)k} - \partial_j \Gamma_{il}^{(\alpha)k} + \Gamma_{iml}^{(\alpha)k} \Gamma_{jl}^{(\alpha)m} - \Gamma_{jml}^{(\alpha)k} \Gamma_{il}^{(\alpha)m} \right) \\ &= \frac{(1-\alpha)^2}{4} \sum_m \sum_{r_1} \sum_{r_2} \left[\left(\partial_i \partial_m \partial_{r_1} \psi g^{r_1 k} \right) \left(\partial_j \partial_l \partial_{r_2} \psi g^{r_2 m} \right) - \left(\partial_j \partial_m \partial_{r_1} \psi g^{r_1 k} \right) \left(\partial_i \partial_l \partial_{r_2} \psi g^{r_2 m} \right) \right] \\ &= \frac{(1-\alpha)^2}{4} \sum_m \sum_{r_1} \sum_{r_2} \left[\left(T_{imr_1} g^{r_1 k} \right) \left(T_{jlr_2} g^{r_2 m} \right) - \left(T_{jmr_1} g^{r_1 k} \right) \left(T_{ilr_2} g^{r_2 m} \right) \right] \\ &= \frac{(1-\alpha)^2}{4} \sum_m \left(T_{im}^k T_{jl}^m - T_{jm}^k T_{il}^m \right). \end{aligned}$$

Tomando $\alpha = 0$, obtemos $R_{lij}^k = R_{lij}^{(0)k}$,

$$R_{lij}^k = \frac{1}{4} \sum_{m \in J} \left(T_{im}^k T_{jl}^m - T_{jm}^k T_{il}^m \right). \quad (4.19)$$

¹⁹ Por vezes, ele é referenciado como o próprio tensor de curvatura de Riemann. Aqui, consideramos o endomorfismo de curvatura de Riemann que usualmente é utilizado na geometria riemanniana como o campo $\binom{3}{1}$ -tensorial dado por

$$\begin{aligned} R : \mathfrak{X}(M) \times \mathfrak{X}(M) \times \mathfrak{X}(M) &\longrightarrow \mathfrak{X}(M) \\ (A, B, C) &\longmapsto R(A, B)C = \nabla_A^{(0)} \nabla_B^{(0)} C - \nabla_B^{(0)} \nabla_A^{(0)} C - \nabla_{[A, B]}^{(0)} C, \end{aligned}$$

que, em um dado sistema de coordenadas, tem coeficientes R_{kij}^l dados por

$$R(\partial_k, \partial_i) \partial_j = \sum_{l \in J} R_{kij}^l \partial_l.$$

²⁰ A fim de tornar a notação mais sucinta, o conjunto sobre o qual estamos somando nos somatórios a seguir foi omitido. Em todos eles, a soma é feita sobre J .

Lembrando que o tensor de curvatura de Riemann pode ser obtido, em coordenadas, através de abaixamento de índice de R_{lij}^k , podemos escrever, no sistema ϑ , sua generalização utilizando α -conexões como

$$\begin{aligned} R_{lijk}^{(\alpha)} &= \sum_{m_1 \in J} g_{km_1} R_{lij}^{(\alpha)m_1} \\ &= \frac{(1-\alpha)^2}{4} \sum_{m_1 \in J} \sum_{m_2 \in J} g_{km_1} \left(T_{im_2}^{m_1} T_{jl}^{m_2} - T_{jm_2}^{m_1} T_{il}^{m_2} \right) \\ &= \frac{(1-\alpha)^2}{4} \sum_{m \in J} \left(T_{ikm} T_{jl}^m - T_{jkm} T_{il}^m \right). \end{aligned}$$

Assim, o tensor de curvatura usual $R_{lijk} = R_{lijk}^{(0)}$ escrito nessas coordenadas é

$$R_{lijk} = \frac{1}{4} \sum_{m \in J} \left(T_{ikm} T_{jl}^m - T_{jkm} T_{il}^m \right).$$

Desse modo, observamos que a maneira como a curvatura varia em diferentes direções depende do tensor 3-simétrico, ou seja, de derivadas de terceira ordem da função ψ . Isso pode ser entendido, também, como uma motivação para chamar ψ de função potencial, uma vez que, de maneira menos formal, podemos compreender o tensor de curvatura de Riemann como dependente de derivadas de segunda ordem de um dado campo de vetores na variedade.

Além disso, como mencionado brevemente em (AY *et al.*, 2017), 4.19 apresenta uma forma que se assemelha àquela das equações de Witten–Dijkgraaf–Verlinde–Verlinde (tipicamente mencionadas como equações WDVV), cujas interpretações geométricas manifestam-se usualmente no estudo das variedades de Frobenius. Como uma rápida explicação acerca dessa declaração, podemos afirmar que elas constituem um sistema de equações diferenciais parciais não lineares cujo estudo iniciou-se mais sistematicamente a partir da década de 1980 (MAGRI, 2016). Como mostra (MAGRI, 2016), a construção dessas equações pode ser feita partindo de uma função $\psi(\vartheta^1, \vartheta^2, \dots, \vartheta^n)$ de n coordenadas $\vartheta = (\vartheta^1, \vartheta^2, \dots, \vartheta^n)$ de M a valores reais. Tomando as componentes $c_i = \frac{\partial \text{Hess}(\psi(\vartheta))}{\partial \vartheta^i}$, com $i \in J$, do gradiente da matriz hessiana $\text{Hess}(\psi(\vartheta))$ de ψ com respeito a ϑ e assumindo invertibilidade para algum c_i , digamos $c_{i'}$, as WVDV equações consistem em equações da forma

$$c_j c_{i'}^{-1} c_l - c_l c_{i'}^{-1} c_j = 0, \quad (4.20)$$

para $j, l \in J$. Na empreitada de fornecer uma compreensão geométrica a essas equações, Boris Dubrovin interpretou $c_{i'}$ como a matriz das componentes de uma métrica pseudo-riemanniana em M escrita em um sistema de coordenadas distinguido²¹ sob a hipótese de que $\frac{\partial c_{i'}}{\partial \vartheta^i} = 0$ para todo $i \in J$. Esse é um dos pressupostos para a definição de espaços onde a geometria deles está intrinsecamente relacionada às equações WDVV, chamados de variedades de Frobenius. Como

²¹ Por distinguido, entende-se que as equações de WDVV, nesse sistema, tomam o formato de 4.20. Isso pois Dubrovin estava interessado em fornecer uma interpretação geométrica intrínseca para essas equações (MAGRI, 2016).

dissertar sobre essas variedades foge à finalidade deste trabalho, para saber mais sobre elas, veja (MAGRI, 2016). Denotando, então, por g_{kr} as componentes da métrica em ϑ sistema de coordenadas utilizado no cálculo de R^l_{kij} , junto a ψ que vinhamos utilizando, temos que 4.20, nesse contexto, escreve-se como

$$\sum_{m \in J} \sum_{r \in J} (\partial_j \partial_i \partial_r \psi g^{rm} \partial_i \partial_m \partial_k \psi - \partial_i \partial_l \partial_r \psi g^{rm} \partial_j \partial_m \partial_k \psi) = 0,$$

ou, em termos do tensor 3-simétrico,

$$\sum_{m \in J} (T_{ikm} T_{jl}^m - T_{jkm} T_{il}^m) = 0. \quad (4.21)$$

Atentando-se a 4.19, vemos que 4.21 é verdadeira apenas se a curvatura R^l_{kij} neste ponto for nula. Desse modo, fica explícita a relação entre essas expressões.

Finalizado, então, esse comentário sobre as equações WDVV, contudo, ainda nos atendo à discussões a respeito de curvatura, podemos afirmar que uma variedade ser plana em relação a uma α -conexão é uma característica que se preserva por dualização, como expressa o seguinte resultado.

Proposição 12. (AY *et al.*, 2017) Seja (M, g, T) variedade estatística. Se $R^{(-\alpha)}$ é nulo, então $R^{(\alpha)}$ também é.

Demonstração. Em um sistema de coordenadas arbitrário, temos que

$$\begin{aligned} \nabla_i^{(-\alpha)} \nabla_j^{(-\alpha)} \partial_k &= \nabla_i^{(-\alpha)} \left(\nabla_j^{(0)} \partial_k + \frac{\alpha}{2} \mathcal{H}(\partial_j, \partial_k) \right) \\ &= \nabla_i^{(-\alpha)} \nabla_j^{(0)} \partial_k + \frac{\alpha}{2} \nabla_i^{(-\alpha)} \mathcal{H}(\partial_j, \partial_k) \\ &= \nabla_i^{(0)} \nabla_j^{(0)} \partial_k + \frac{\alpha}{2} \mathcal{H}(\partial_i, \nabla_j^{(0)} \partial_k) + \frac{\alpha}{2} \nabla_i^{(0)} \mathcal{H}(\partial_j, \partial_k) + \frac{\alpha^2}{4} \mathcal{H}(\partial_i, \mathcal{H}(\partial_j, \partial_k)), \end{aligned}$$

Por simples permutações de índices e trocas de sinais, podemos obter expressões análogas para $\nabla_j^{(-\alpha)} \nabla_i^{(-\alpha)} \partial_k$, $\nabla_i^{(\alpha)} \nabla_j^{(\alpha)} \partial_k$ e $\nabla_i^{(\alpha)} \nabla_j^{(\alpha)} \partial_k$. Observe que

$$\nabla_{[\partial_i, \partial_j]}^{(-\alpha)} \partial_k = \nabla_{[\partial_i, \partial_j]}^{(0)} \partial_k + \frac{\alpha}{2} \mathcal{H}([\partial_i, \partial_j], \partial_k) = 0 = \nabla_{[\partial_i, \partial_j]}^{(0)} \partial_k - \frac{\alpha}{2} \mathcal{H}([\partial_i, \partial_j], \partial_k) = \nabla_{[\partial_i, \partial_j]}^{(\alpha)} \partial_k.$$

Substituindo isso junto a $\nabla_i^{(-\alpha)} \nabla_j^{(-\alpha)} \partial_k$ e $\nabla_j^{(-\alpha)} \nabla_i^{(-\alpha)} \partial_k$ em $R^{(-\alpha)}(\partial_i, \partial_j) \partial_k = 0$, obtemos

$$A = -\frac{\alpha}{2} \left[\mathcal{H}(\nabla_j^{(0)} \partial_k, \partial_i) + \nabla_i^{(0)} \mathcal{H}(\partial_j, \partial_k) - \mathcal{H}(\nabla_i^{(0)} \partial_k, \partial_j) - \nabla_j^{(0)} \mathcal{H}(\partial_i, \partial_k) \right],$$

onde

$$A := \nabla_i^{(0)} \nabla_j^{(0)} \partial_k + \frac{\alpha^2}{4} \mathcal{H}(\partial_i, \mathcal{H}(\partial_j, \partial_k)) - \nabla_j^{(0)} \nabla_i^{(0)} \partial_k - \frac{\alpha^2}{4} \mathcal{H}(\partial_j, \mathcal{H}(\partial_i, \partial_k)).$$

Calculando $R^{(\alpha)}(\partial_i, \partial_j)$ por meio de $\nabla_i^{(\alpha)} \nabla_j^{(\alpha)} \partial_k$ e $\nabla_j^{(\alpha)} \nabla_i^{(\alpha)} \partial_k$, podemos notar que A aparece nesse cálculo, de maneira que é válida a seguinte sequência de igualdades:

$$\begin{aligned} R^{(\alpha)}(\partial_i, \partial_j) \partial_k &= A - \frac{\alpha}{2} \left[\mathcal{H}(\nabla_j^{(0)} \partial_k, \partial_i) - \nabla_i^{(0)} \mathcal{H}(\partial_j, \partial_k) + \mathcal{H}(\nabla_i^{(0)} \partial_k, \partial_j) + \nabla_j^{(0)} \mathcal{H}(\partial_i, \partial_k) \right] \\ &= A - A \\ &= 0, \end{aligned}$$

para todo $i, j, k \in J$. □

Dito isso, voltemo-nos agora à conexão $\nabla^{(-1)}$, que, segundo 4.18, é dada, em ϑ , através dos símbolos

$$\Gamma_{ijk}^{(-1)} = \partial_i \partial_j \partial_k \psi.$$

Vamos definir, então, o sistema de coordenadas $\eta = (\eta_1, \eta_2, \dots, \eta_n)$ em M dado por

$$\eta_i := \partial_i \psi,$$

para todo $i \in J$. Consequentemente, g_{ij} fica escrita, em termos de η , como $g_{ij} = \partial_i \eta_j$. Aplicando, então, a transformada de Legendre

$$\varphi(\eta) = \max_{\vartheta} \left(\sum_{i \in J} \vartheta^i \eta_i - \psi(\vartheta) \right),$$

obtemos que a função potencial de η é dada através de

$$\psi(\vartheta) + \varphi(\eta) - \sum_{i \in J} \vartheta^i \eta_i = 0. \quad (4.22)$$

Com efeito,

$$\partial^i \varphi(\eta) = \frac{\partial \varphi(\eta)}{\partial \eta_i} = \frac{\partial}{\partial \eta_i} \left(\sum_{i \in J} \vartheta^i \eta_i - \psi(\vartheta) \right) = \vartheta^i$$

e, além disso, lembrando que

$$\sum_{j \in J} g_{ij} g^{kj} = \delta_i^k,$$

como, para $i \neq k$,

$$\sum_{j \in J} \frac{\partial \eta_j}{\partial \vartheta^i} \frac{\partial \vartheta^i}{\partial \eta_j} = \frac{\partial \vartheta^i}{\partial \vartheta^i} = 1, \quad \sum_{j \in J} \frac{\partial \eta_j}{\partial \vartheta^i} \frac{\partial \vartheta^k}{\partial \eta_j} = \frac{\partial \vartheta^k}{\partial \vartheta^i} = 0$$

fica claro que

$$g^{ij} = \frac{\partial \vartheta^j}{\partial \eta_i} = \partial^i \partial^j \varphi(\eta).$$

Esse fato, como esperado, está em consonância com 5, de tal modo que tomar como ponto de partida uma função potencial ou basear-se em uma métrica já estabelecida na variedade são abordagens que se equivalem. Esse é um resultado apresentado em (AY *et al.*, 2017) como o seguinte teorema.

Teorema 6. (AY *et al.*, 2017) Uma estrutura dualmente plana, isto é, uma métrica riemanniana g juntamente com duas conexões planas ∇ e ∇^* que são duais em relação a g , é localmente equivalente à informação²² de uma única função convexa ψ , onde a convexidade aqui se refere às coordenadas locais ϑ e não a qualquer métrica.

Por fim, é importante pontuar que, como os resultados apresentados possuem, em sua maioria, um caráter local, as condições de compatibilidade entre sistemas de coordenadas precisam ser satisfeitas com o intuito de produzir uma estrutura que se estenda para toda a variedade. Contudo, como propõe (AY *et al.*, 2017) ao introduzir essa questão, variedades de distribuições de probabilidade não apresentam, em geral, fenômenos singulares, sendo bem comportadas dentro do que foi apresentado (AY *et al.*, 2017). Desse modo, essa discussão perde seu caráter emergencial dentro deste tratamento amplo. Obviamente, uma análise cuidadosa acerca das sobreposições de cartas pode ser feita, entretanto, a geometria da informação, ao menos na generalidade de sua bibliografia, considera esse formalismo um ponto marginal, que usualmente é omitido. Esse comentário coloca-se como uma recomendação sobre aplicar resultados de geometria da informação a espaços que apresentam certas anomalias do ponto de vista geométrico.

²² O autor utiliza o termo “*datum*” que, durante toda a obra, aparece com o sentido de “informação”, como em “*additional datum*”. Por ser um termo impreciso no enunciado do teorema e cuja tradução não foi feita de maneira literal, optamos por fazer essa observação.

APLICAÇÃO: GEOMETRIA DA INFORMAÇÃO E LIMITES QUÂNTICOS DE VELOCIDADE

5.1 A introdução da geometria da informação na busca por limites quânticos de velocidade

Como foi mencionado diversas vezes durante este trabalho, a geometria da informação coloca-se com um potencial de aplicabilidade que se estende para além da matemática. Neste capítulo, apresentaremos uma dessas aplicações, observando como essa área pode ser utilizada no estudo de limites quânticos de velocidade, referidos como QSL, através de uma abordagem geométrica. Para tal, iremos nos basear no artigo de revisão ([PIRES *et al.*, 2016](#)), seguindo fortemente o conteúdo exposto nele.

Dito isso e dialogando com o início de [2](#), podemos destacar que, no contexto de mecânica quântica, existem diversas relações de incerteza consagradas, em que uma das mais célebres diz respeito ao Princípio de Incerteza de Heisenberg. Assim, dados A e B dois observáveis referentes a um sistema quântico, sejam, respectivamente, ΔA e ΔB as raízes quadradas dos desvios quadrados médios de suas medidas em um estado arbitrário $|\psi\rangle$, isto é,

$$\Delta C = \sqrt{\langle \psi | (C - \langle C \rangle)^2 | \psi \rangle}, \quad \langle C \rangle = \langle \psi | C | \psi \rangle$$

para $C = A, B$. Então, é válida a relação

$$\Delta A \Delta B \geq \frac{1}{2} |\langle \psi | [A, B] | \psi \rangle|,$$

chamada de um relação de incerteza entre A e B e onde $[A, B]$ é o comutador de A e B ([PICASSO, 2015](#)), podendo ser entendida como uma desigualdade que envolve as incertezas nas medições

dos dois observáveis. Desse modo, é legítimo o questionamento acerca de uma relação dessa forma para o tempo e a energia. Esse questionamento encontra como primeiro obstáculo o fato do tempo não ser um operador associado a um observável, mas um parâmetro universal relativo à dinâmica do sistema. Isso conduziu a uma falta de consenso entre físicos sobre interpretar a não comutabilidade entre o operador energia $i\hbar \frac{\partial}{\partial t}$ e o tempo t como uma impossibilidade de os medir simultaneamente¹ (PRICE; CHISSICK; HEISENBERG, 1977). Nesse contexto de busca por uma relação de incerteza envolvendo essas grandezas, interpretou-se tal relação como uma desigualdade que expressa o limite mínimo imposto pela física sobre o tempo de evolução de um sistema entre dois estados quânticos distinguíveis (PIRES *et al.*, 2016). Durante o desenvolvimento histórico dessa questão, houveram diversos resultados. Um deles, atribuído aos físicos soviéticos Leonid Mandelstam e Igor Tamm, diz respeito a dinâmicas unitárias² geradas por um hamiltoniano dependente do tempo e que evolui os sistemas de um estado puro $|\psi(0)\rangle$ para outro estado puro $|\psi(\tau)\rangle$, sendo esses estados distinguíveis (PIRES *et al.*, 2016). Nesse caso,

$$\tau \geq \frac{\hbar \arccos(|\langle \psi(\tau) | \psi(0) \rangle|)}{\Delta E}. \quad (5.1)$$

Outro resultado, obtido pelo engenheiro canadense-estadunidense Norman Margolus e pelo físico russo-estadunidense Lev Levitin, é válido para sistemas fechados cujo hamiltoniano H é independente do tempo e evolui os sistemas entre dois estados puros ortogonais (PIRES *et al.*, 2016). Nesse contexto, o resultado de Margolus e Levitin pode ser traduzido na relação

$$\tau \geq \frac{\hbar \pi}{2E}, \quad E := \langle H \rangle. \quad (5.2)$$

Desse modo, unindo 5.1 a 5.2, obteve-se um limite mais apertado³ para τ quando há o tratamento de uma dinâmica unitária que evolui um estado puro $|\psi(0)\rangle$ a outro $|\psi(\tau)\rangle$ ortogonal ao primeiro (PIRES *et al.*, 2016). Nesse caso, como $|\langle \psi(\tau) | \psi(0) \rangle| = 0$, temos que

$$\tau \geq \max \left\{ \frac{\hbar \pi}{2\Delta E}, \frac{\hbar \pi}{2E} \right\}.$$

Posto isso, diante dessa busca por QSLs, aspectos geométricos podem conduzir a tais limites, de maneira que a geometria da informação apresenta-se como colaborativa nesse processo. Isso se dá por meio do estudo de distinguibilidade entre dois estados de um sistema através do uso

¹ É importante pontuar que não há consenso acerca de simultaneidade de medições envolvendo t e outras grandezas físicas, sendo um termo comumente usado com um sentido amplo e intuitivo (PRICE; CHISSICK; HEISENBERG, 1977).

² Dado um espaço vetorial complexo V munido de um produto interno $\langle \cdot | \cdot \rangle$, um operador U em V é chamado de unitário em V se ele preserva a norma de todos os vetores de V (WAN, 2019). Assim, uma dinâmica é unitária se ela é regida por operadores unitários, como ficará mais claro quando apresentarmos a decomposição de uma evolução através de operadores de Kraus.

³ Por “limite mais apertado” entende-se um limitante melhor para τ ; mais rígido. Esse é um termo de difícil tradução do inglês para o português e alguns autores optam por manter a adjetivação “tighter” para caracterizar limitantes mais apertados para τ . Aqui, optamos por o traduzir como “apertado” e fazer essa observação.

de modelos estatísticos, uma vez que o conjunto de operadores densidade sobre o espaço de Hilbert de um sistema possui estrutura intrínseca de variedade diferenciável (PIRES *et al.*, 2016). Então, por meio da introdução de uma métrica riemanniana nesse conjunto, o comprimento das geodésicas com respeito a sua conexão de Levi-Civita pode ser utilizado para gerar um QSL. Assim, a busca por esses limitantes se traduz, primeiramente, na procura pela métrica riemanniana que melhor expressa a física dessas dinâmicas quânticas. Perceba que é apropriado assumir que essa métrica é contrativa sob mapas completamente positivos que preservam traço e a justificativa para isso é simplesmente porque tais mapas mantêm as características dos operadores densidade (PIRES *et al.*, 2016). No contexto clássico, assumir isso consiste em considerar que a métrica é contrativa sob mapas estocásticos. Tratando de geometria sobre espaços de distribuições de probabilidade de dimensão finita, como consequência do Teorema de Caracterização de Chentsov, temos que a métrica de Fisher é a única métrica riemanniana contrativa sob tais mapas. Já no contexto quântico, a métrica de Fisher coloca-se como constitutiva de uma família de infinitas métricas que satisfazem essa condição de contração (PIRES *et al.*, 2016). É por meio de análises geométricas dessa família que se obtém alguns QSLs. Apresentemos, então, um processo para a obtenção desses limites.

5.2 A obtenção de QSLs geométricos

Sabemos que qualquer sistema quântico está associado a um espaço de Hilbert, usualmente denotado por \mathcal{H} . Dito isso, definamos o espaço de operadores densidade sobre esse espaço.

Definição 30 (Operador densidade). (PIRES *et al.*, 2016) Seja \mathcal{H} um espaço de Hilbert. Um operador sobre \mathcal{H} que é hermitiano, positivo semidefinido e de traço unitário é chamado de um operador densidade sobre \mathcal{H} . Denotamos o espaço de operadores densidade sobre \mathcal{H} por $\mathcal{D}(\mathcal{H})$.

Assim, os estados de \mathcal{H} são representados por $\mathcal{D}(\mathcal{H})$ munida de uma métrica riemanniana $\langle \cdot | \cdot \rangle$. Como vimos em 5.1, métricas que são contrativas apresentam importância destacada no estudo de QSLs geométricos, de maneira que é coerente defini-las.

Definição 31 (Métricas contrativas). (PIRES *et al.*, 2016) Seja $F : \mathcal{H} \rightarrow \mathcal{H}$ um mapa completamente positivo e que preserva traço. Uma métrica riemanniana g em $\mathcal{D}(\mathcal{H})$ é dita contrativa sob F se, para todo $\rho_1, \rho_2 \in \mathcal{D}(\mathcal{H})$,

$$\mathcal{L}(F(\rho_1), F(\rho_2)) \leq \mathcal{L}(\rho_1, \rho_2),$$

onde \mathcal{L} expressa a distância geodésica induzida por⁴ g .

⁴ Perceba que essa definição pressupõe que $(\mathcal{D}(\mathcal{H}), g)$ é geodesicamente completa, suposição que transpassa todo o artigo (PIRES *et al.*, 2016).

Como argumenta (PIRES *et al.*, 2016), no caso de dimensão finita, as métricas contrativas⁵ são caracterizadas pelas chamadas funções de Morozova-Chentsov, ou seja, para cada uma dessas funções, existe uma métrica riemanniana em $\mathcal{D}(\mathcal{H})$ que satisfaz a condição de contração em questão.

Definição 32 (Funções de Morozova-Chentsov). (PIRES *et al.*, 2016) Uma função $f : (0, +\infty) \rightarrow (0, +\infty)$ é chamada uma função de Morozova-Chentsov se ela satisfaz as condições abaixo:

1. Monotonicidade para operadores: para quaisquer operadores A e B semipositivos definidos tais que $A \leq B$, então $f(A) \leq f(B)$.
2. Autoinversibilidade: para todo $t \in (0, +\infty)$, f satisfaz $f(t) = tf(1/t)$.
3. Normalização: $f(1) = 1$.

É importante notar que essas funções são limitadas inferiormente e superiormente, de forma respectiva, pelas funções de Morozova-Chentsov f_{\min} e f_{\max} dadas por

$$f_{\min} = \frac{2t}{1+t}, \quad f_{\max} = \frac{1+t}{2}.$$

Em outras palavras, dada f uma função de Morozova-Chentsov, para todo $t \in (0, +\infty)$, $f_{\min}(t) \leq f(t) \leq f_{\max}(t)$ (PIRES *et al.*, 2016).

Ademais, denotamos por g^f a métrica contrativa associada à função de Morozova-Chentsov f . Assim, dada g^f , baseando-se no trabalho (PETZ, 1996) do físico húngaro Dénes Petz sobre as descobertas de Chentsov e da física russa Elena Morozova, (PIRES *et al.*, 2016) apresenta o elemento quadrático de comprimento ds^2 entre dois operadores densidade vizinhos ρ e $\rho + d\rho$ induzido por g^f , a menos de um fator constante, como

$$ds^2 = \frac{1}{4} \left[\sum_{j \in J} \frac{(d\rho_{jj})^2}{p_j} + 2 \sum_{\substack{j, l \in J \\ j < l}} c^f(p_j, p_l) |d\rho_{jl}|^2 \right], \quad (5.3)$$

onde ρ foi escrita em sua decomposição espectral $\rho = \sum_{j \in J} p_j |j\rangle\langle j|$, com $0 < p_j \leq 1$ para todo $j \in J$ e $\sum_{j \in J} p_j = 1$, onde $d\rho_{jl} := \langle j | d\rho | l \rangle$ e onde c^f é a função dada por

$$c^f(p_j, p_l) := \frac{1}{p_l f(p_j/p_l)}. \quad (5.4)$$

Observe que o primeiro termo de 5.3 já foi apresentado, correspondendo à métrica de Fisher clássica⁶ em ρ . Já o segundo termo é responsável pela não unicidade da métrica, sendo dependente

⁵ A partir de agora, a especificação da contração ser sob mapas completamente positivos e que preservam traço será omitida.

⁶ A caracterização “clássica” advém da existência de uma segunda métrica, conhecida como métrica de Fisher quântica. Além disso, é comum dizer que essa expressão para a métrica de Fisher é em ρ , diferentemente do caso de modelos estatísticos fora do contexto quântico, em que se costuma dizer que a métrica é no parâmetro j , tal qual fizemos em 2.2.

das funções de Morozova-Chentsov e sendo uma expressão da coerência⁷ de $d\rho$ com respeito à base de autovetores de ρ (PIRES *et al.*, 2016).

Note que, como g^f é contrativa, então seus comprimentos geodésicos são medidas de distinguibilidade genuínas para densidades de probabilidade (PIRES *et al.*, 2016). Isso pode ser explicado do ponto de vista de sistemas abertos⁸. Em um sistema quântico aberto, a informação é perdida para o ambiente, de maneira que, na média, estados inicialmente bem distinguíveis tornam-se cada vez mais parecidos a partir da perspectiva das medições quânticas. Explicado isso, por mais que os comprimentos geodésicos, nesse contexto, sejam medidas de distinguibilidade genuínas, o cálculo de geodésicas e, conseqüentemente, de seus comprimentos, pode ser um trabalho complicado, como sabemos e já comentamos em 3.5. O caso da família de métricas g^f em $\mathcal{D}(\mathcal{H})$ não é diferente. Como destaca (PIRES *et al.*, 2016), as expressões analíticas para as distâncias geodésicas, nesse caso, são conhecidas apenas para $g^{f_{\max}}$ e $g^{f_{WY}}$, onde

$$f_{WY}(t) = \frac{(\sqrt{t} + 1)^2}{4}.$$

No caso de $g^{f_{\max}}$, essa métrica é conhecida como métrica de Fisher quântica, uma vez que ela apresenta um formato análogo ao da métrica de Fisher clássica, sendo definida, em um estado quântico parametrizado ρ_ξ , por meio da matriz de informação de Fisher quântica de entradas

$$F_{ij}(\xi) := \frac{1}{2} \text{Tr}(\rho_\xi \{L_i, L_j\}),$$

onde $\{L_i, L_j\}$ indica o anticomutador de L_i e L_j e onde L_i e L_j são chamados de derivadas logarítmicas simétricas, dadas por

$$\partial_i \rho_\xi = \frac{1}{2} (\rho_\xi L_i - L_i \rho_\xi)$$

para todo $i \in J$ (LIU *et al.*, 2019). Desse modo, a menos de um fator multiplicativo, o elemento quadrático ds^2 dessa métrica escreve-se como

$$ds^2 = \frac{1}{2} \left[\sum_{j,l \in J} \frac{|\langle j | d\rho | l \rangle|^2}{p_j + p_l} \right] = \frac{1}{4} \left[\sum_{j \in J} \frac{(d\rho_{jj})^2}{p_j} + 2 \sum_{\substack{j,l \in J \\ j < l}} \frac{2}{p_j + p_l} |d\rho_{jl}|^2 \right],$$

que está de acordo com 5.3, e a distância geodésica $\mathcal{L}^{QF}(\rho_1, \rho_2)$ entre $\rho_1, \rho_2 \in \mathcal{D}(\mathcal{H})$ induzida por $g^{f_{\max}}$ é dada por

$$\mathcal{L}^{QF}(\rho_1, \rho_2) = \arccos \left(\sqrt{F(\rho_1, \rho_2)} \right), \quad (5.5)$$

onde $F(\rho_1, \rho_2) := (\text{Tr}[\sqrt{\sqrt{\rho_1} \rho_2 \sqrt{\rho_1}}])^2$ é conhecida como fidelidade de Uhlmann entre ρ_1 e ρ_2 (PIRES *et al.*, 2016). Já, quando a função de Morozova-Chentsov é f_{WY} , $g^{f_{WY}}$ recebe o nome

⁷ Um estado é coerente em uma dada base do espaço se os seus termos não diagonais, nessa base, são não nulos.

⁸ Um sistema aberto é um sistema físico bem delimitado tal que é possível haver fluxo de matéria, trabalho e energia, por si mesma, através desse limite (BASU, 2018).

de métrica de Wigner-Yanase, cuja distância geodésica entre ρ_1 e $\rho_2 \in \mathcal{D}(\mathcal{H})$, nesse caso, escreve-se como

$$\mathcal{L}^{\text{WY}}(\rho_1, \rho_2) = \arccos(A(\rho_1, \rho_2)), \quad (5.6)$$

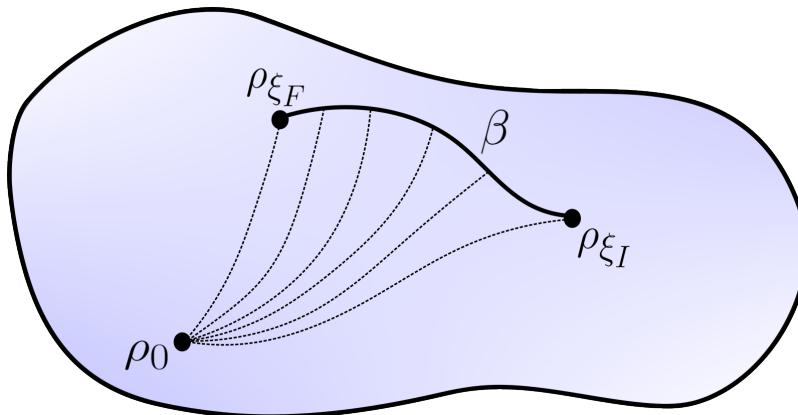
onde $A(\rho_1, \rho_2) = \text{Tr}(\sqrt{\rho_1}\sqrt{\rho_2})$ é chamada de afinidade quântica entre ρ_1 e ρ_2 e, assim como a fidelidade de Uhlmann, é uma medida de distinguibilidade entre estados quânticos (PIRES *et al.*, 2016).

Outrossim, mesmo sendo complexo o cálculo de tais distâncias geodésicas, podemos estudar a evolução de um estado quântico genérico a fim de, por meio da característica de minimização das geodésicas, expressar QSLs. Assim, dado $\rho_0 \in \mathcal{D}(\mathcal{H})$, a evolução dinâmica de ρ_0 pode ser descrita por meio dos chamados operadores de Kraus K_j^ξ , com $j \in J$, que constituem uma família de operadores $\{K_j^\xi\}_{j \in J}$ que satisfazem $\sum_{j \in J} K_j^{\xi\dagger} K_j^\xi = 1$ e dependem do r -parâmetro $\xi = \{\xi_1, \xi_2, \dots, \xi_r\}$ que define a dinâmica (PIRES *et al.*, 2016). Desse modo, a evolução de ρ_0 determinada por ξ pode ser escrita como

$$\rho_\xi = \sum_{j \in J} K_j^\xi \rho_0 K_j^{\xi\dagger},$$

de maneira que ela pode ser interpretada como um caminho⁹ em $\mathcal{D}(\mathcal{H})$ iniciando em ρ_0 . Considere, então, uma dinâmica evolutiva ρ_ξ em que ξ varia analiticamente de um parâmetro ξ_I para um ξ_F , determinando, assim, uma curva β diferenciável ligando ρ_{ξ_I} a ρ_{ξ_F}

Figura 13 – Representação gráfica de uma variação analítica na evolução dinâmica de ρ_0 . As curvas tracejadas indicam evoluções intermediárias entre ρ_{ξ_I} e ρ_{ξ_F} , que determinam a curva β ligando ρ_{ξ_I} a ρ_{ξ_F}



Fonte: Adaptada de Pires *et al.* (2016, p 021031-3).

Sabemos que o comprimento ℓ_β^f de β depende da escolha de g^f e, para cada f fixado, esse comprimento é dado por

$$\ell_\beta^f = \int_\beta ds.$$

⁹ A notação ρ_ξ indica tanto o ponto final da evolução de ρ_0 determinada por ξ quanto essa própria dinâmica evolutiva.

Perceba que, como a variação analítica do parâmetro ξ foi tomada arbitrariamente, o comprimento de β é limitado inferiormente pelo comprimento da geodésica minimizante ligando ρ_{ξ_l} a ρ_{ξ_F} .

Ademais, note que, como a evolução ρ_ξ é função de ξ , temos que

$$d\rho_\xi = \sum_{k=1}^r \partial_k \rho_\xi d\xi_k.$$

Escrevendo ρ_ξ em sua decomposição espectral como $\rho_\xi = \sum_{j \in J} p_j |j\rangle\langle j|$, com $0 < p_j \leq 1$ para todo $j \in J$ e $\sum_{j \in J} p_j = 1$, temos, pela regra do produto para derivadas, que

$$\partial_k \rho_\xi = \sum_{j \in J} \{ (\partial_k p_j) |j\rangle\langle j| + p_j [(\partial_k |j\rangle)\langle j| + |j\rangle(\partial_k \langle j|)] \}.$$

Usando a identidade $(\partial_k \langle j|) |l\rangle = -\langle j| \partial_k |l\rangle$ para todo $j, l \in J$ e $k \in R := \{1, 2, \dots, r\}$, obtemos a expressão

$$\begin{aligned} \langle j | \partial_k \rho_\xi | l \rangle &= \sum_{m \in J} [\langle j | \partial_k p_m | m \rangle \langle m | l \rangle + \langle j | p_m (\partial_k | m \rangle) \langle m | l \rangle + \langle j | m \rangle p_m (\partial_k \langle m |) | l \rangle] \\ &= \partial_k p_j \delta_{jl} + (p_l - p_j) \langle j | \partial_k | l \rangle. \end{aligned}$$

Consequentemente, temos que, para todo $j, l \in J$ e $k \in R$, é válida a seguinte sequência de igualdades

$$\begin{aligned} \langle j | d\rho_\xi | l \rangle &= \sum_{k \in R} \langle j | \partial_k \rho_\xi d\xi_k | l \rangle \\ &= \sum_{k \in R} [\partial_k p_j \delta_{jl} + i(p_l - p_j) i \langle j | \partial_k | l \rangle] d\xi_k \\ &= \sum_{k \in R} [\partial_k p_j \delta_{jl} + i(p_l - p_j) \mathcal{A}_{jl}^k] d\xi_k, \end{aligned}$$

onde $\mathcal{A}_{jl}^k := i \langle j | \partial_k | l \rangle$. Desse modo,

$$|\langle j | d\rho_\xi | l \rangle|^2 = \begin{cases} \sum_{k_1, k_2 \in R} (\partial_{k_1} p_j) (\partial_{k_2} p_j) d\xi_{k_1} d\xi_{k_2}, & \text{se } i = j \\ \sum_{k_1, k_2 \in R} (p_j - p_l)^2 \mathcal{A}_{jl}^{k_1} \mathcal{A}_{lj}^{k_2} d\xi_{k_1} d\xi_{k_2}, & \text{caso contrário.} \end{cases}$$

Com isso, podemos escrever ds^2 em termos de $d\xi_k$, com $k \in R$. De fato,

$$\begin{aligned} ds^2 &= \frac{1}{4} \left[\sum_{j \in J} \sum_{k_1, k_2 \in R} \frac{(\partial_{k_1} p_j) (\partial_{k_2} p_j)}{p_j} + 2 \sum_{\substack{j, l \in J \\ j < l}} \sum_{k_1, k_2 \in R} c^f(p_j, p_l) (p_j - p_l)^2 \mathcal{A}_{jl}^{k_1} \mathcal{A}_{lj}^{k_2} \right] d\xi_{k_1} d\xi_{k_2} \\ &= \sum_{k_1, k_2 \in R} \left[\frac{1}{4} \sum_{j \in J} \frac{(\partial_{k_1} p_j) (\partial_{k_2} p_j)}{p_j} + \frac{1}{2} \sum_{\substack{j, l \in J \\ j < l}} c^f(p_j, p_l) (p_j - p_l)^2 \mathcal{A}_{jl}^{k_1} \mathcal{A}_{lj}^{k_2} \right] d\xi_{k_1} d\xi_{k_2} \\ &= \sum_{k_1, k_2 \in R} g_{k_1, k_2}^f d\xi_{k_1} d\xi_{k_2}, \end{aligned}$$

onde

$$g_{k_1 k_2}^f := \mathcal{F}_{k_1 k_2} + \mathcal{Q}_{k_1 k_2}^f,$$

com

$$\mathcal{F}_{k_1 k_2} := \frac{1}{4} \sum_{j \in J} \frac{(\partial_{k_1} p_j)(\partial_{k_2} p_j)}{p_j}, \quad (5.7)$$

$$\mathcal{Q}_{k_1 k_2}^f := \frac{1}{2} \sum_{\substack{j, l \in J \\ j < l}} c^f(p_j, p_l) (p_j - p_l)^2 \mathcal{A}_{jl}^{k_1} \mathcal{A}_{lj}^{k_2}. \quad (5.8)$$

Comparando essa expressão com 5.3, observamos que o termo $\mathcal{F}_{k_1 k_2}$ é uma expressão comum à toda família de métricas g^f , enquanto que $\mathcal{Q}_{k_1 k_2}^f$ refere-se à contribuição advinda da função de Morozova-Chentsov f .

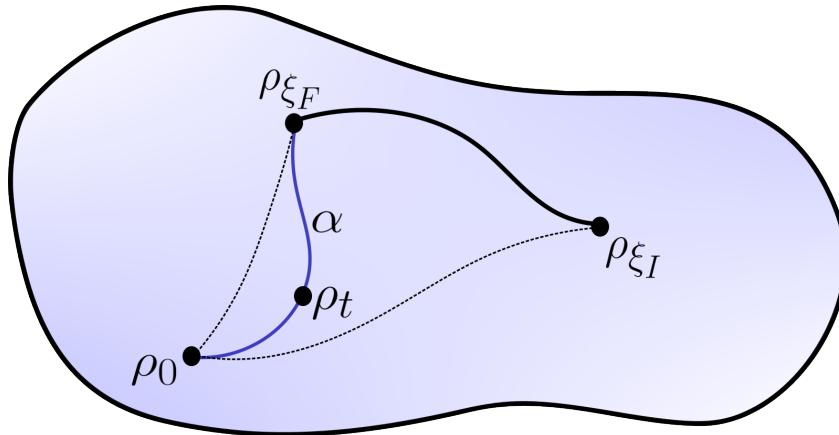
Considere, agora, que ξ é função do tempo. Sendo τ o tempo de evolução de ρ_0 , tome a parametrização $t \in [0, \tau] \mapsto \xi(t)$ tal que $\xi(0) = \xi_I$ e $\xi(\tau) = \xi_F$. Logo, o comprimento da geodésica minimizante ligando ρ_0 ao estado final ρ_τ é um limitante inferior para o comprimento da curva α de evolução do estado ρ_0 a ρ_τ , ou seja,

$$\mathcal{L}^f(\rho_0, \rho_\tau) \leq \ell_\alpha^f(\rho_0, \rho_\tau), \quad (5.9)$$

onde

$$\ell_\alpha^f(\rho_0, \rho_\tau) = \int_0^\tau \sqrt{\sum_{k_1, k_2 \in R} g_{k_1 k_2}^f \frac{d\xi_{k_1}}{dt} \frac{d\xi_{k_2}}{dt}} dt.$$

Figura 14 – Representação gráfica da evolução dinâmica de ρ_0 no tempo. Em azul, temos o estado evolutivo de ρ_0 no tempo $t \in [0, \tau]$. Inicialmente, a curva se aproxima daquela que fornece a evolução ρ_{ξ_I} e, conforme t chega perto de τ , a curva azul torna-se próxima de ρ_{ξ_F} .



Fonte: Adaptada de Pires *et al.* (2016, p 021031-3).

Note que, para cada função de Morozova-Chentsov f , 5.9 fornece um QSL geométrico. Assim, fixada a dinâmica α que evolui ρ_0 , a fim de mensurarmos quão apertado é o QSL

fornecido por uma métrica g^f , definamos a diferença abaixo relativa à geodésica minimizante ligando ρ_0 a ρ_τ :

$$\delta_\alpha^f := \frac{\ell_\alpha^f(\rho_0, \rho_\tau) - \mathcal{L}^f(\rho_0, \rho_\tau)}{\mathcal{L}^f(\rho_0, \rho_\tau)}.$$

Assim, para cada dinâmica α , o QSL geométrico mais apertado obtido a partir de 5.9 é aquele referente a f que minimiza δ_α^f (PIRES *et al.*, 2016). Logo, o QSL geométrico mais apertado é

$$\mathcal{L}^{f_M}(\rho_0, \rho_\tau) \leq \ell_\alpha^{f_M}(\rho_0, \rho_\tau), \quad (5.10)$$

onde f_M é a função de Morozova-Chentsov tal que $f_M = \arg \min_f \delta_\alpha^f$ (PIRES *et al.*, 2016). Perceba que, por mais que 5.10 seja apertado quando comparado com 5.9, ainda há a dificuldade de calcular o lado esquerdo dessa desigualdade. Sendo assim, a fim de ilustrar tais QSLs, exploremos três exemplos nas próximas seções deste capítulo, analisando os casos em que esse cálculo pode ser realizado sem grandes dificuldades. Nesse ponto, finalizamos as aplicações do que foi apresentado, especialmente no capítulo 2, à busca por QSLs. Por completude, apresentaremos os exemplos seguintes a fim de ilustrar o papel do análogo quântico da métrica de Fisher no contexto de limites quânticos de velocidade.

5.3 Dinâmica unitária

Considere um sistema quântico fechado com estado inicial ρ_0 que evolui por uma dinâmica unitária. Como a soma de operadores unitários não necessariamente é um operador unitário, a decomposição de uma evolução unitária ρ_ξ em operadores de Kraus escreve-se como $\rho_\xi = U_\xi \rho_0 U_\xi^\dagger$, onde U_ξ é unitário. Considerando tal evolução, como transformações unitárias não alteram pureza¹⁰, a matriz de ρ_ξ não depende de ξ na base dos seus autovetores e, consequentemente, $\partial_k p_j = 0$ para todo $i \in J$ e para todo $k \in R$, onde p_j são os autovalores de ρ_ξ (PIRES *et al.*, 2016). Assim, observando 5.5, para todo $k_1, k_2 \in R$, temos que $\mathcal{F}_{k_1 k_2} = 0$. Deste modo, no caso unitário, $g_{k_1 k_2}^f$ simplifica-se em $\mathcal{Q}_{k_1 k_2}$. No caso de dinâmicas unitárias, é conhecido que $\mathcal{A}_{jl}^k = \frac{1}{\hbar} \langle j | \Delta H_k^\xi | l \rangle$, onde $\Delta H_k^\xi = H_k^\xi - \langle H_k^\xi \rangle$, com $\langle H_k^\xi \rangle = \text{Tr}(\rho_\xi H_k^\xi)$ e $H_k^\xi = -i\hbar U_\xi \partial_k U_\xi^\dagger$ (PIRES *et al.*, 2016). Nesse caso,

$$g_{k_1 k_2}^f = \mathcal{Q}_{k_1 k_2} = \frac{1}{2\hbar^2} \sum_{\substack{j, l \in J \\ j < l}} c^f(p_j, p_l) (p_j - p_l)^2 \langle j | \Delta H_{k_1}^\xi | l \rangle \langle l | \Delta H_{k_2}^\xi | j \rangle, \quad (5.11)$$

que é uma expressão dependente da escolha da função de Morozova-Chentsov. Uma vez que as geodésicas de $g^{f_{\max}}$ são conhecidas, estudemos tal caso. Como

$$c^{f_{\max}}(p_j, p_l) = \frac{2}{p_j + p_l},$$

¹⁰ A pureza de $\rho \in \mathcal{D}(\mathcal{H})$ é definida como $\text{Tr}(\rho^2)$.

obtemos que 5.11 escreve-se como

$$g_{k_1 k_2}^{QF} = \frac{1}{2\hbar^2} \sum_{j,l \in J} \frac{(p_j - p_l)^2}{p_j + p_l} \langle j | \Delta H_{k_1}^\xi | l \rangle \langle l | \Delta H_{k_2}^\xi | j \rangle.$$

Uma vez que $0 < p_m \leq 1$ para todo $m \in J$, sabemos que vale a desigualdade

$$\frac{(p_j - p_l)^2}{p_j + p_l} \leq p_j + p_l.$$

Desse modo, obtemos um limitante superior para $g_{k_1 k_2}^{QF}$, dado por

$$g_{k_1 k_2}^{QF} \leq \frac{1}{2\hbar^2} \sum_{j,l \in J} (p_j + p_l) \langle j | \Delta H_{k_1}^\xi | l \rangle \langle l | \Delta H_{k_2}^\xi | j \rangle = \frac{\mathcal{C}(\Delta H_{k_1}^\xi, \Delta H_{k_2}^\xi)}{\hbar^2},$$

onde

$$\mathcal{C}(\Delta H_{k_1}^\xi, \Delta H_{k_2}^\xi) := \sum_{j,l \in J} (p_j + p_l) \langle j | \Delta H_{k_1}^\xi | l \rangle \langle l | \Delta H_{k_2}^\xi | j \rangle = \frac{1}{2} \text{Tr} \left[\rho_\xi \{ \Delta H_{k_1}^\xi, \Delta H_{k_2}^\xi \} \right]$$

é chamada de covariância simetrizada de $\Delta H_{k_1}^\xi$ e $\Delta H_{k_2}^\xi$ (PIRES *et al.*, 2016). Assim, substituindo isso em 5.9, obtemos a relação

$$\begin{aligned} \mathcal{L}^{QF}(\rho_0, \rho_\tau) &\leq \int_0^\tau \sqrt{\sum_{k_1, k_2 \in R} g_{k_1 k_2}^f \frac{d\xi_{k_1}}{dt} \frac{d\xi_{k_2}}{dt}} dt \\ &\leq \frac{1}{\hbar} \int_0^\tau \sqrt{\sum_{k_1, k_2 \in R} \mathcal{C}(\Delta H_{k_1}^\xi, \Delta H_{k_2}^\xi) \frac{d\xi_{k_1}}{dt} \frac{d\xi_{k_2}}{dt}} dt. \end{aligned}$$

Por simplicidade, suponha que ξ seja um 1-parâmetro, digamos $\xi = t$. Nesse caso, H_k^ξ reduz-se apenas a $H_t = -i\hbar U_t \partial_t U_t^\dagger$ e $\mathcal{C}(\Delta H_{k_1}^\xi, \Delta H_{k_2}^\xi)$ corresponde à variância quadrática de H_t . Com efeito,

$$\begin{aligned} \mathcal{C}(\Delta H_t, \Delta H_t) &= \frac{1}{2} \sum_{j,l \in J} (p_j + p_l) \langle j | \Delta H_t | l \rangle \langle l | \Delta H_t | j \rangle \\ &= \text{Tr} \left[\rho_\xi (\Delta H_t)^2 \right] \\ &= \text{Tr} \left\{ \rho_\xi \left[(H_t)^2 - H_t \langle H_t \rangle - \langle H_t \rangle H_t + \langle H_t \rangle^2 \right] \right\} \\ &= \left\langle (H_t)^2 \right\rangle - \langle H_t \rangle^2. \end{aligned}$$

Nesse caso, então, o limite que havíamos obtido escreve-se como

$$\tau^{-1} \mathcal{L}^{QF}(\rho_0, \rho_\tau) \leq \frac{1}{\hbar \tau} \int_0^\tau \sqrt{\langle H_t^2 \rangle - \langle H_t \rangle^2} dt = \frac{1}{\hbar} \overline{\Delta E},$$

onde

$$\overline{\Delta E} := \frac{1}{\tau} \int_0^\tau \sqrt{\langle H_t^2 \rangle - \langle H_t \rangle^2} dt$$

é a variância média de H_t (PIRES *et al.*, 2016). Consequentemente, temos o QSL

$$\tau \geq \frac{\hbar}{\Delta E} \mathcal{L}^{\text{QF}}(\rho_0, \rho_\tau),$$

que possui um formato próximo àquele obtido por Mandelstam e Tamm (PIRES *et al.*, 2016). Isso se deve ao fato de eles terem estudado estados puros distinguíveis, uma vez que a fidelidade de Uhlmann é nula nesse caso e, como consequência de 5.5, observamos que $\mathcal{L}^{\text{QF}}(\rho_0, \rho_\tau) = \frac{\pi}{2}$.

Ademais, assim como no caso de $g^{f_{\max}}$, a expressão das geodésicas de 5.11 são conhecidas quando a função de Morozova-Chentsov escolhida é f_{WY} . Desse modo, estudemos 5.9 quando g^f é a métrica de Wigner-Yanase. Introduzindo f_{WY} na expressão 5.4, temos que

$$c^f(p_j, p_l) = \frac{4}{(\sqrt{p_j} + \sqrt{p_l})}.$$

Substituindo em 5.11, obtemos

$$\begin{aligned} g_{k_1 k_2}^{\text{WY}} &= \frac{2}{\hbar^2} \sum_{\substack{j, l \in J \\ j < l}} \left(\frac{p_j - p_l}{\sqrt{p_j} + \sqrt{p_l}} \right)^2 \langle j | \Delta H_{k_1}^\xi | l \rangle \langle l | \Delta H_{k_2}^\xi | j \rangle \\ &= -\frac{1}{\hbar^2} \text{Tr} \left(\left[\sqrt{\rho}, \Delta H_{k_1}^\xi \right] \left[\sqrt{\rho}, \Delta H_{k_2}^\xi \right] \right) \\ &= \frac{2}{\hbar^2} \mathcal{G} \left(\Delta H_{k_1}^\xi, \Delta H_{k_2}^\xi \right), \end{aligned}$$

onde $\mathcal{G} := -\frac{1}{2} \text{Tr} \left(\left[\sqrt{\rho}, \Delta H_{k_1}^\xi \right] \left[\sqrt{\rho}, \Delta H_{k_2}^\xi \right] \right)$ (PIRES *et al.*, 2016). Substituindo na desigualdade 5.9, obtemos

$$\mathcal{L}^{\text{WY}}(\rho_0, \rho_\tau) \leq \frac{\sqrt{2}}{\hbar} \int_0^\tau \sqrt{\sum_{k_1, k_2 \in R} \mathcal{G} \left(\Delta H_{k_1}^\xi, \Delta H_{k_2}^\xi \right) \frac{d\xi_{k_1}}{dt} \frac{d\xi_{k_2}}{dt}} dt.$$

Tal como foi feito para o caso da métrica de Fisher quântica, vamos supor que ξ é o 1-parâmetro $\xi = t$, de maneira que H_k^ξ escreve-se como $H_t = -i\hbar U_t \partial_t U_t^\dagger$. Nesse caso, \mathcal{G} reduz-se a chamada informação assimétrica $\mathcal{I}(\rho_t, H_t) := -\frac{1}{2} \text{Tr} \left(\left[\sqrt{\rho}, \Delta H_t \right]^2 \right)$ entre o estado evoluído ρ_t e o H_t . Então, nesse caso, 5.9 escreve-se como

$$\tau^{-1} \mathcal{L}^{\text{WY}}(\rho_0, \rho_\tau) \leq \frac{\sqrt{2}}{\hbar} \sqrt{\mathcal{I}},$$

onde

$$\sqrt{\mathcal{I}} := \frac{1}{\tau} \int_0^\tau \sqrt{\mathcal{I}(\rho_t, H_t)} dt$$

é a raiz quadrada da informação assimétrica média entre ρ_t e H_t (PIRES *et al.*, 2016). Logo, temos o QSL geométrico

$$\tau \geq \frac{\hbar}{\sqrt{2} \sqrt{\mathcal{I}}} \mathcal{L}^{\text{WY}}(\rho_0, \rho_\tau).$$

Usando o fato de $\mathcal{I}(\rho_t, H_t) \leq \langle (H_t)^2 \rangle - \langle H_t \rangle^2$, observamos que

$$\tau \geq \frac{\hbar}{\sqrt{2} \sqrt{\mathcal{I}}} \mathcal{L}^{\text{WY}}(\rho_0, \rho_\tau) \geq \frac{1}{\sqrt{2}} \frac{\hbar}{\Delta E} \mathcal{L}^{\text{WY}}(\rho_0, \rho_\tau).$$

Suponha, então, que ρ_0 e ρ_τ comutam. Nesse caso, $\sqrt{\mathcal{F}(\rho_0, \rho_\tau)} = A(\rho_0, \rho_\tau)$ e, consequentemente,

$$\tau \geq \frac{\hbar}{\Delta E} \mathcal{L}^{QF}(\rho_0, \rho_\tau) \geq \frac{\hbar}{\sqrt{2\Delta E}} \mathcal{L}^{WY}(\rho_0, \rho_\tau),$$

de maneira que, em termos da variância média de H_t , a métrica de Fisher quântica é mais apertada do que a métrica de Wigner-Yanase nesse contexto (PIRES *et al.*, 2016). Esse é um resultado advindo não apenas da comutatividade da distribuição inicial ρ_0 e da final ρ_τ , mas também da dinâmica ser unitária. Estudemos, agora, o que ocorre em dois casos em que essa unitariedade da dinâmica é retirada e o sistema deixa de ser fechado.

5.4 Dinâmica não unitária - canal de defasagem paralela

Estudando o caso de um *qubit*¹¹, um estado na esfera de Bloch¹² pode ser denotado por

$$\rho = \frac{1}{2}(1 + \vec{r} \cdot \vec{\sigma}),$$

onde $\vec{r} = (r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta)$, com $r \in [0, 1]$, $\theta \in [0, \pi]$ e $\phi \in [0, 2\pi)$, é chamado de vetor de Bloch e $\vec{\sigma}$ é o vetor das matrizes de Pauli¹³ (PIRES *et al.*, 2016). Considere, então, a equação

$$\frac{\partial \rho(t)}{\partial t} = H(\rho) + L(\rho)$$

que rege sistemas quânticos abertos, chamada de equação mestra de Lindblad, onde $H(\rho) := -i[H, \rho]$ representa a evolução unitária dada pelo hamiltoniano H e $L(\rho)$ é um termo liouvilliano, que descreve o ruído no sistema quântico (PIRES *et al.*, 2016). Sendo assim, considere

$$H = \frac{\omega_0}{2} \sigma_3, \quad L(\rho) = -\frac{\Gamma}{2} (\rho - \sigma_3 \rho \sigma_3), \quad (5.12)$$

onde ω_0 é a frequência unitária e Γ é a taxa de decoerência¹⁴ na base computacional. É importante citar que, por frequência unitária, entende-se a frequência que surge quando se evolui o estado pela equação de Schrödinger, sem o termo liouvilliano. Dito isso, sob as considerações de 5.12, o canal que introduz o ruído no sistema é chamado de canal de defasagem paralela, uma vez que $L(\rho)$ tem a mesma direção dada pela matriz de Pauli que H .

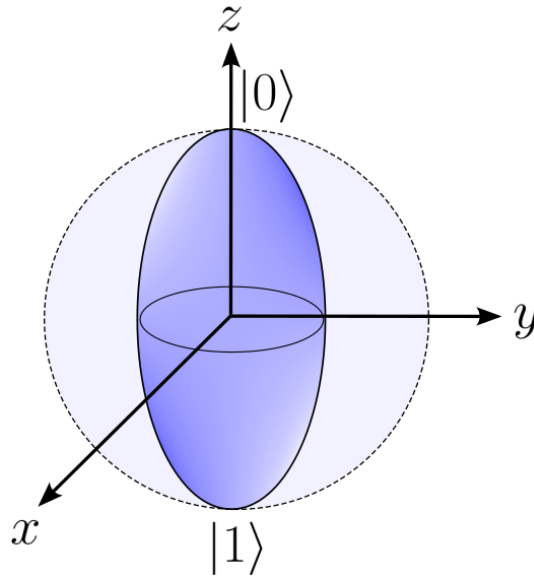
¹¹ Um *qubit* é um sistema quântico de dois níveis distintos. Ele é um bit que possui dois estados distinguíveis $|0\rangle$ e $|1\rangle$, podendo existir em um estado de superposição deles (WATERLOO, 2022).

¹² Grosso modo, a esfera de Bloch é uma representação geométrica esférica de um *qubit*. Os pontos na casca da esfera correspondem a estados puros, enquanto que os pontos do interior referem-se a misturas de estados. Os polos norte e sul representam, respectivamente, $|0\rangle$ e $|1\rangle$.

¹³ Para saber mais, veja o capítulo 2 de (NIELSEN; CHUANG, 2010)

¹⁴ Uma decoerência é um processo para eliminar os termos não diagonais de um estado coerente em uma base. No caso de *qubits*, a coerência é usualmente expressa em relação à base computacional $\{|0\rangle, |1\rangle\}$.

Figura 15 – Representação gráfica da esfera de Bloch no caso de um canal de defasagem paralela. O canal contrai a esfera mantendo a direção z intacta, assim como o centro da esfera deformada, fato que fornece a esse canal a adjetivação unital



Fonte: Elaborada pelo autor.

Nesse caso, a evolução de ρ_0 , descrita em termos de operadores de Kraus, é da forma

$$\rho_t = K_0 \rho_0 K_0^\dagger + K_1 \rho_0 K_1^\dagger,$$

onde

$$K_0 = \sqrt{q_+} \begin{pmatrix} e^{-i\omega_0 t/2} & 0 \\ 0 & e^{i\omega_0 t/2} \end{pmatrix}, \quad K_1 = \sqrt{q_-} \begin{pmatrix} e^{-i\omega_0 t/2} & 0 \\ 0 & -e^{i\omega_0 t/2} \end{pmatrix},$$

com $q_\pm := \frac{1 \pm q_t}{2}$ para $q_t := e^{-\Gamma t}$ (PIRES *et al.*, 2016). Tomando ρ_t em sua decomposição espectral, obtemos

$$\rho_t = \sum_{j=\pm} p_j |\theta_t, \phi_t\rangle_j \langle \theta_t, \phi_t|_j,$$

onde

$$p_\pm := \frac{1}{2} (1 \pm r_0 \xi_t),$$

$$\xi_t := \sqrt{\cos^2 \theta_0 + q_t^2 \sin^2 \theta_0},$$

$$|\theta_t, \phi_t\rangle_\pm = \frac{1}{N_\pm} \left[(\cos \theta_0 \pm \xi_t) |0\rangle + e^{i(\omega_0 t + \phi_0)} q_t \sin \theta_0 |1\rangle \right]$$

para N_\pm constantes de normalização e r_0, θ_0, ϕ_0 coordenadas do estado inicial na esfera de Bloch (PIRES *et al.*, 2016). Note que

$$\partial_t p_\pm = \pm \frac{r_0 q_t \left(\frac{dq_t}{dt} \right) \sin^2 \theta_0}{2 \xi_t}.$$

Substituindo isso em 5.7, temos que

$$\begin{aligned}\mathcal{F} &= \frac{1}{4} \left[\frac{(\partial_t p_+)(\partial_t p_+)}{p_+} + \frac{(\partial_t p_-)(\partial_t p_-)}{p_-} \right] \\ &= \frac{1}{4} \left[\frac{r_0^2 q_t^2 \left(\frac{dq_t}{dt}\right)^2 \sin^4 \theta_0}{4\xi_t^2 \left(\frac{1}{2} + \frac{1}{2} r_0 \xi_t\right)} + \frac{r_0^2 q_t^2 \left(\frac{dq_t}{dt}\right)^2 \sin^4 \theta_0}{4\xi_t^2 \left(\frac{1}{2} - \frac{1}{2} r_0 \xi_t\right)} \right] \\ &= \frac{r_0^2 q_t^2 \left(\frac{dq_t}{dt}\right)^2 \sin^4 \theta_0}{4\xi_t^2 (1 - r_0^2 \xi_t^2)}.\end{aligned}$$

Além disso, perceba que

$$\pm \langle \theta_t, \phi_t | \partial_t | \theta_t, \phi_t \rangle_{\mp} = \frac{q_t \sin^2 \theta_0}{N_+ N_-} \left[i\omega_0 q_t \mp \left(\frac{dq_t}{dt} \right) \frac{\cos \theta_0}{\xi_t} \right],$$

de maneira que a expressão 5.8 escreve-se como

$$\mathcal{Q}^f = \frac{1}{8} \left[\omega_0^2 q_t^2 + \frac{\cos^2 \theta_0 \left(\frac{dq_t}{dt}\right)^2}{\xi_t^2} \right] r_0^2 \sin^2 \theta_0 c^f(p_+, p_-),$$

onde as constantes de normalização foram desconsideradas. Com esses resultados, temos que os termos \mathcal{F} e \mathcal{Q}^f da métrica g^f anulam-se para todos os estados que, inicialmente, estão sobre o eixo z , independente da função de Morozova-Chentsov f . Isso se deve ao fato dos estados nesse eixo não serem afetados pelo canal de defasagem, como é possível observar em 15. Estudando, então, a qualidade das métricas de Fisher quântica e de Wigner-Yanase na obtenção de QSLs, os autores de (PIRES *et al.*, 2016) identificaram um fenômeno que destoava daquele observado para o caso unitário. Em regimes de baixa frequência, ou seja, conforme ω_0 aproxima-se de 0, a métrica g^{WY} produz limites mais apertados que g^{QF} , especialmente para raios iniciais r_0 próximos da superfície da esfera e para ângulos iniciais θ_0 perto de 0 ou π . Em outras palavras, nesse regime, a métrica de Wigner-Yanase produz QSLs tão melhores que os obtidos pela métrica de Fisher quântica quanto os estados iniciais estejam próximos dos elementos da base computacional na esfera. Já, conforme a frequência aumenta, esse comportamento se inverte e, para estados iniciais nas proximidades de $|0\rangle$ e $|1\rangle$, g^{QF} produz limites mais apertados que g^{WY} , como expõe (PIRES *et al.*, 2016). Isso mostra que, em sistemas mais realísticos, nem sempre a métrica de Fisher quântica é a melhor no estudo de QSLs. Seguindo o que foi proposto em (PIRES *et al.*, 2016), vejamos se esse é um comportamento que se estende para canais que deformam a esfera ao longo do eixo z , também.

5.5 Dinâmica não unitária - canal de amortecimento de amplitude

Ainda considerando um *qubit*, a evolução

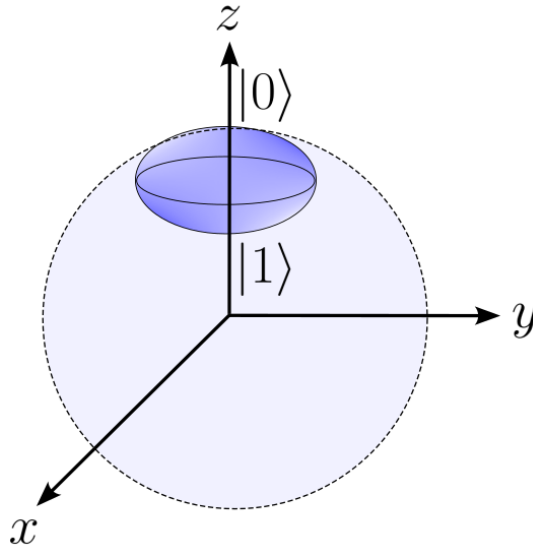
$$\rho_t = K_0 \rho_0 K_0^\dagger + K_1 \rho_0 K_1^\dagger,$$

de ρ_0 é descrita em termos de operadores de Kraus

$$K_0 = \begin{pmatrix} 1 & 0 \\ 0 & \sqrt{1-\xi_t} \end{pmatrix}, \quad K_1 = \begin{pmatrix} 0 & \sqrt{\xi_t} \\ 0 & 0 \end{pmatrix},$$

onde $\xi_t := 1 - e^{-\Gamma t}$ (PIRES *et al.*, 2016). Esse canal desloca o centro da esfera de Bloch na direção positiva do eixo z , deformando a esfera em direção ao estado $|0\rangle$.

Figura 16 – Representação gráfica da esfera de Bloch no caso de um canal de amortecimento de amplitude. Ele desloca a esfera na direção z junto ao seu centro, sendo um canal não unital



Fonte: Elaborada pelo autor.

Nesse caso, a decomposição espectral da evolução ρ_t do estado inicial ρ_0 é dada por

$$\rho_t = \sum_{j=\pm} p_j |\theta_t, \phi_t\rangle_j \langle \theta_t, \phi_t|_j,$$

onde

$$p_{\pm} := \frac{1}{2}(1 \pm \vartheta_t),$$

$$\vartheta_t := \sqrt{1 - \zeta_t(1 - \xi_t)},$$

$$\zeta_t := 1 - r_0^2 + \xi_t(1 - r_0 \cos \theta_0)^2,$$

$$\varsigma_t := \xi_t + r_0(1 - \xi_t) \cos \theta_0,$$

$$|\theta_t, \phi_t\rangle_{\pm} = \frac{1}{\sqrt{N_{\pm}}} \left[(\varsigma_t \pm \vartheta_t) |0\rangle + e^{i\phi_0} r_0 \sqrt{1 - \xi_t} \sin \theta_0 |1\rangle \right],$$

e N_{\pm} são constantes de normalização (PIRES *et al.*, 2016). Perceba que

$$\begin{aligned} \partial_t p_{\pm} &= \pm \frac{1}{2} \partial_t \vartheta_t \\ &= \pm \frac{\Gamma e^{-\Gamma t}}{4 \vartheta_t} \left[(\xi_t - 1)(1 - r_0 \cos \theta_0)^2 + \zeta_t \right]. \end{aligned}$$

Substituindo em 5.7, obtemos

$$\begin{aligned}\mathcal{F} &= \frac{1}{4} \left[\frac{(\partial_+ p_+)(\partial_+ p_+)}{p_+} + \frac{(\partial_t p_-)(\partial_t p_-)}{p_-} \right] \\ &= \frac{\Gamma^2 e^{-2\Gamma t}}{32 \vartheta_t^2} \left[(\xi_t - 1)(1 - r_0 \cos \theta_0)^2 + \zeta_t \right]^2 \left(\frac{1 + \vartheta_t + 1 - \vartheta_t}{1 - \vartheta_t^2} \right). \\ &= \frac{\Gamma^2 e^{-2\Gamma t}}{16 \vartheta_t^2 \zeta_t (1 - \xi_t)} \left[(\xi_t - 1)(1 - r_0 \cos \theta_0)^2 + \zeta_t \right]^2\end{aligned}$$

Além disso, note que

$$\begin{aligned}\pm \langle \theta_t, \phi_t | \partial_t | \theta_t, \phi_t \rangle_{\mp} &= \frac{\Gamma e^{-\Gamma t}}{\sqrt{N_+ N_-}} \left\{ \left[(\zeta_t \pm \vartheta_t) r_0 \cos \theta_0 - 1 \right. \right. \\ &\quad \left. \left. \mp \frac{1}{2 \vartheta_t} \left((1 - r_0 \cos \theta_0)^2 (1 - \xi_t) - \zeta_t \right) \right] + \frac{r_0^2 \sin^2 \theta_0}{2} \right\}.\end{aligned}$$

Adicionando essa expressão em 5.8, obtemos

$$\mathcal{Q}^f = \frac{r_0^2 \sin^2 \theta_0 (2 - \zeta_t)^2 c^f(p_+, p_-)}{32 \vartheta_t^2 (1 - \xi_t)},$$

onde N_{\pm} e o termo de fase global $\Gamma e^{-\Gamma t}$ foram suprimidos. Então, assim como no caso do canal de defasagem paralela, quando o ângulo inicial θ_0 é 0 ou π , a contribuição de \mathcal{Q}^f desaparece, de maneira que os comprimentos geodésicos são independentes da escolha da função de Morozova-Chentsov. Enquanto que \mathcal{F} anula-se no caso de defasagem paralela quando $\theta_0 \in \{0, \pi\}$, aqui ele rege a geometria advinda de g^f .

Ademais, tal como foi citado na seção 5.4, os autores de (PIRES *et al.*, 2016) também compararam, nesse caso, quão apertados são os QSLs produzidos pela métrica de Fisher quântica em relação àqueles obtidos pela métrica de Wigner-Yanase. Seus resultados mostram que os QSLs obtidos por meio de g^{WY} são quase saturados em toda a esfera de Bloch, ou seja, $\delta^{WY} \approx 0$ para quase todo ρ_0 na esfera. Já g^{QF} quase se satura nas proximidades de $|0\rangle$ e, conforme ρ_0 afasta-se desse ponto, observa-se que $\delta^{QF} \geq \delta^{WY}$. Esse comportamento expressa o fato de $|0\rangle$ ser invariante pelo canal de amortecimento de amplitude (PIRES *et al.*, 2016). Assim, diferentemente dos demais dois casos abordados neste capítulo, a métrica de Wigner-Yanase produz QSLs mais apertados que a métrica de Fisher quântica em quase toda a esfera de Bloch.

CONCLUSÃO

Este trabalho configura-se como uma apresentação sobre a geometria da informação, partindo de modelos estatísticos e seguindo até as α -conexões, exibindo a aplicabilidade das noções dessa área no cálculo de limites quânticos de velocidade. Sendo assim, ele organiza-se como um compilado de conceitos e resultados que sustentam esse campo de estudos, não havendo, desse modo, um único resultado que se sobressai como a conclusão desta dissertação, mas sim um conjunto deles.

Dito isso, sendo a geometria da informação uma área que se conecta com diversas outras, em especial com a estatística, seu caráter interdisciplinar transpassa seu desenvolvimento histórico, de maneira que os modelos estatísticos, possuindo uma estrutura intrínseca de variedade diferenciável, podem ser munidos da métrica de Fisher, do Tensor de Amari-Chentsov e de uma família uniparamétrica de conexões lineares, essa última sendo a família das α -conexões. Isso permite o estudo de dissimilaridade entre as distribuições de probabilidades desses modelos por meio de geodésicas as ligando, geodésicas essas que podem ser selecionadas de acordo com o modelo, escolhendo um fator α em $[-1, 1]$. Desse modo, a α -geometria permite uma ampliação das análises estatísticas em modelos estatísticos.

Além disso, por um lado, na perspectiva estatística advinda do Teorema do Limite Central, uma distribuição de probabilidades se sobressai como uma distribuição de convergência no contexto de grandes amostras e em que há independência e igual distribuição de probabilidade. Essa distribuição é a gaussiana. Por outro, a geometria hiperbólica destaca-se como uma das três geometrias de curvatura constante para superfícies. Esses dois objetos, importantes cada qual para sua respectiva área, conectam-se na geometria da informação por meio da métrica de Fisher g , de maneira que ela, no modelo gaussiano, escreve-se, no sistema de coordenadas natural dele, como $g = 2g_{\mathbb{H}^2}$, onde $g_{\mathbb{H}^2}$ é a métrica hiperbólica no modelo do Plano de Lobachevsky escrita em coordenadas retangulares. Dessa forma, pudemos concluir que a geometria advinda da métrica de Fisher no modelo gaussiano é, essencialmente, hiperbólica.

Além disso, partindo de uma abordagem extrínseca para a geometria da informação finita, mergulhando os espaços de parâmetros no espaço de medidas positivas $\mathcal{M}_+(\mathcal{X})$ sobre o conjunto finito não vazio $\mathcal{X} = \{x_1, x_2, \dots, x_{n+1}\}$, observamos, primeiramente, que o espaço de distribuições de probabilidade positivas $\mathcal{P}_+(\mathcal{X})$ sobre \mathcal{X} , munido da métrica de Fisher g induzida de $\mathcal{M}_+(\mathcal{X})$, é isométrico ao setor esférico positivo

$$S_{2,+}(\mathcal{X}) := \left\{ f \in \mathcal{F}(\mathcal{X}) \mid f(x_i) > 0, i \in I, \sum_{i \in I} f(x_i)^2 = 4 \right\} \subset \mathcal{F}(\mathcal{X})$$

equipado com a métrica herdada da métrica euclidiana em $\mathcal{F}(\mathcal{X})$, de maneira que podemos concluir que $(\mathcal{P}_+(\mathcal{X}), g)$ possui curvatura constante positiva.

Ademais, por meio do Teorema de Caracterização de Chentsov e de sua extensão para o tensor de Amari-Chentsov em $\mathcal{P}_+(\mathcal{X})$, pudemos concluir que a métrica de Fisher e que esse campo tensorial são os únicos campos, respectivamente, 2- e 3-covariantes não triviais em $\mathcal{P}_+(\mathcal{X})$ que são preservados por morfismos de Markov induzidos por núcleos de Markov congruentes, sendo esse um resultado que não se aplica a campos n -covariantes, com $n \geq 4$.

Outrossim, sendo o tensor de Amari-Chentsov definido através de duas conexões lineares em $\mathcal{M}_+(\mathcal{X})$, essas conexões apresentam relativa importância na geometria da informação. Chamadas de conexão mistura e conexão exponencial, suas geodésicas, ao menos no último caso, expressam a nomenclatura dada a elas, uma vez que as geodésicas no caso da exponencial são dadas por

$$\begin{aligned} \widehat{\gamma}^{(e)} : \mathbb{R} &\longrightarrow \mathcal{R}(\mathcal{X}) \\ t &\longmapsto \frac{\exp\left(t \frac{da}{d\eta}\right)}{N} \eta. \end{aligned}$$

onde $\mathcal{R}(\mathcal{X})$ é notação tanto para $\mathcal{M}_+(\mathcal{X})$ quanto para $\mathcal{P}_+(\mathcal{X})$ e onde

$$N = \begin{cases} 1, & \text{se } \mathcal{R}(\mathcal{X}) = \mathcal{M}_+(\mathcal{X}), \\ \xi\left(\exp\left(t \frac{da}{d\eta}\right)\right), & \text{se } \mathcal{R}(\mathcal{X}) = \mathcal{P}_+(\mathcal{X}). \end{cases}$$

Já as geodésicas no caso da mistura são da forma

$$\begin{aligned} \widehat{\gamma}^{(m)} : (t^-, t^+) &\longrightarrow \mathcal{R}(\mathcal{X}) \\ t &\longmapsto \eta + ta \end{aligned}$$

com

$$t^- := -\min \left\{ \frac{\eta_i}{a_i} \mid i \in I, a_i > 0 \right\}, \quad t^+ := \min \left\{ \frac{\eta_i}{|a_i|} \mid i \in I, a_i < 0 \right\}$$

sob a convenção $\min \emptyset = +\infty$. Essas duas conexões são um caso destacado das α -conexões. Assim, esse mesmo cálculo pode ser feito para se obter as α -geodésicas. Em $\mathcal{M}_+(\mathcal{X})$, obtemos que elas são da forma

$$\begin{aligned} \gamma^{(\alpha)} : (t^-, t^+) &\longrightarrow \mathcal{M}_+(\mathcal{X}) \\ t &\longmapsto \left(\mu^{\frac{1-\alpha}{2}} + t \frac{(1-\alpha)}{2} \mu^{-\frac{1+\alpha}{2}} a \right)^{\frac{2}{1-\alpha}}, \end{aligned}$$

onde

$$t^- := -\min \left\{ \frac{2\mu_i}{(1-\alpha)a_i} \mid i \in I, a_i > 0 \right\}, \quad t^+ := \min \left\{ \frac{2\mu_i}{(1-\alpha)|a_i|} \mid i \in I, a_i < 0 \right\}$$

e onde $\gamma^{(\alpha)}(t) \rightarrow \mu e^{at/\mu}$ quando $\alpha \rightarrow 1^-$. Já as α -geodésicas em $\mathcal{P}_+(\mathcal{X})$ são de difícil determinação, uma vez que a equação da geodésica, nesse caso, consiste em um sistema de equações diferenciais não lineares de segunda ordem com dependências não lineares entre si e com $\alpha \in [-1, 1]$ arbitrário.

Ademais, retornando a uma abordagem intrínseca e apresentadas as noções de variedade dualística e variedade estatística, exibimos a prova de que uma estrutura estatística em uma variedade diferenciável induz uma estrutura dualística livre de torção e vice-versa. Além disso, estudando as famílias de distribuições exponenciais e mistura, pudemos concluir que elas podem ser modeladas por variedades hessianas, uma vez que $\nabla^{(1)}$ é plana na primeira assim como $\nabla^{(-1)}$ é plana na segunda e, além disso, a métrica de Fisher nelas escreve-se nos sistemas de coordenadas afim respectivos a cada conexão como $g_{ij} = \text{Hess}(f)$, onde f , no caso das distribuições mistura, consiste na entropia de Shannon dessas distribuições.

Além disso, estudando estruturas dualmente planas (g, ∇, ∇^*) em uma variedade diferenciável M , observamos que os sistemas de coordenadas afins para ∇ e para ∇^* conectam-se através da transformada de Legendre. Então, expusemos a prova de que a nulidade do tensor de curvatura de Riemann é uma característica mantida por dualização, ou seja, $R^{(-\alpha)} = 0$ implica que $R^{(\alpha)} = 0$. Além disso, explicamos que, por um lado, dada (g, ∇, ∇^*) uma estrutura dualmente plana em M e dado um sistema de coordenadas afim para cada uma das conexões, é possível encontrar duas funções estritamente convexas $\psi, \varphi : M \rightarrow \mathbb{R}$ tais que g e g^{-1} escrevem-se como a hessiana dessa funções em relação aos sistemas afim. Já, por outro lado, mostramos que a recíproca também é verdadeira, de maneira que finalizamos notando que uma estrutura dualmente plana em M é equivalente, ao menos localmente, à informação de uma única função convexa, com a convexidade tomada em relação ao sistema afim para ∇ .

Finalmente, observamos a aplicabilidade da geometria da informação na teoria de informação quântica, em que ela é utilizada na procura por limites quânticos de velocidade geométricos. Isso é feito através do estudo de distinguibilidade entre dois estados de um sistema por meio do uso de modelos estatísticos. Munindo esses modelos com métricas contrativas sob mapas semipositivos e que preservam traço, métricas essas associadas univocamente à funções de Morozova-Chentsov, destacamos que os elementos de comprimento quadrático induzidos por elas entre dois operadores densidade vizinhos escrevem-se fazendo uso da métrica de Fisher clássica. Analisando, então, a qualidade dos QSLs produzidos tanto pela métrica de Fisher quântica quanto pela métrica de Wigner-Yanase, pontuamos que a primeira produz limites mais apertados em relação à segunda no caso de dinâmicas unitárias em que o estado inicial comuta com o final. Quando essa dinâmica deixa de ser unitária, no caso de *qubits*, destacamos que essa relação hierárquica não é mais válida, sendo que, no caso de um canal de amortecimento de

amplitude, a métrica de Wigner-Yanase satura-se em quase toda a esfera de Bloch, mostrando, nesse caso, sua superioridade em relação à métrica de Fisher quântica para a obtenção de QSLs geométricos.

REFERÊNCIAS

- AKIN, E. **The Geometry of Population Genetics**. 3. ed. [S.l.]: Springer, 1979. v. 31. (Lecture notes in Biomathematics, v. 31). ISBN 978-3540097112. Citado na página 23.
- AMARI, S. i. Information geometry and its applications: Convex function and dually flat manifold. In: NIELSEN, F. (Ed.). **Emerging Trends in Visual Computing: LIX Fall Colloquium, ETVC 2008, Palaiseau, France, November 18-20, 2008. Revised Invited Papers**. Alemanha: Springer, 2009. p. 75–102. ISBN 978-3-642-00826-9. Acesso em 18-04-2023. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-642-00826-9_4>. Citado na página 98.
- AMARI, S. i. **Information geometry and its applications**. 1. ed. [S.l.]: Springer Tokyo, 2016. v. 194. (Applied Mathematical Sciences, v. 194). ISBN 978-4-431-55978-8. Citado na página 67.
- AMARI, S. i.; KURATA, K.; NAGAOKA, H. Information geometry of boltzmann machines. **IEEE Transactions on Neural Networks**, v. 3, n. 2, p. 260–271, 1992. Acesso em 28-04-2023. Disponível em: <<https://ieeexplore.ieee.org/document/125867/>>. Citado na página 20.
- AMARI, S. i.; NAGAOKA, H. **Methods of Information Geometry**. Estados Unidos da América: American Mathematical Society, 2000. v. 191. (Translations of Mathematical Monographs, v. 191). ISBN 0-8218-0531-2. Citado nas páginas 20, 23, 24, 25, 34, 35, 38, 41, 42 e 66.
- ARORA, J. S. **Introduction to Optimum Design**. 3. ed. Estados Unidos da América: Academic Press, 2011. ISBN 978-0123813756. Citado na página 43.
- ASH, R. B. **Information Theory**. [S.l.]: Dover Books on Mathematics, 1990. ISBN 978-0486665214. Citado na página 19.
- AY, N.; JOST, J.; LÊ, H. V.; SCHWACHHÖFER, L. **Information geometry**. Suíça: Springer, 2017. v. 64. (Ergebnisse der Mathematik und ihrer Grenzgebiete, v. 64). ISBN 978-3-319-56477-7. Citado nas páginas 20, 21, 22, 23, 24, 25, 26, 27, 28, 34, 44, 45, 46, 51, 55, 56, 58, 59, 60, 66, 67, 69, 71, 73, 74, 75, 76, 77, 78, 79, 81, 86, 88, 89, 90, 91, 92, 93, 94, 96, 103, 104, 105, 106 e 133.
- BAKHVALOV, N. S.; BOROVKOV, A. A.; DOBRUSHIN, R. L.; ZABRODIN, A. V.; ZOLOTAREV, V. M.; IBRAGIMOV, I. A.; PROKHOROV, Y. V.; SEVAST'YANOV, B. A.; SINAI, Y. G.; SKOROKHOD, A. V.; STATULEVICHUS, V. A.; KHAS'MINSKII, R. Z.; KHOLEVO, A. S.; CHIBISOV, D. M.; SHIRYAEV, A. N. Memorial: Nikolai nikolaevitch chentsov. **Theory of Probability & Its Applications**, v. 38, n. 3, p. 506–515, 1994. Acesso em 27-04-2023. Disponível em: <<https://epubs.siam.org/doi/10.1137/1138047>>. Citado na página 21.
- BARTLE, R. G. **The elements of integration and Lebesgue measure**. Estados Unidos da América: Wiley-Interscience, 1995. ISBN 0-471-04222-6. Citado nas páginas 31 e 52.
- BASU, D. K. **Dictionary of Material Science and High Energy Physics**. Estados Unidos da América: CRC Press, 2018. (Comprehensive dictionary of physics). ISBN 9781420049855. Citado na página 111.

- BELL, G. **The Basics of Selection**. [S.l.]: Springer-Science+Business Media, B.V., 1997. ISBN 978-0412055317. Citado na página 23.
- BLOM, G. **Probability and Statistics: Theory and Applications**. 1. ed. Estados Unidos da América: Springer-Vedag, 1989. ISBN 978-1461281580. Citado na página 36.
- CAMPBELL, L. L. An extended čencov characterization of the information metric. In: **Proceedings of the American Mathematical Society**. Estados Unidos da América: American Mathematical Society, 1986. v. 98, n. 1, p. 135–141. Acesso em 28-04-2023. Disponível em: <<https://www.ams.org/journals/proc/1986-098-01/S0002-9939-1986-0848890-5/>>. Citado nas páginas 21 e 52.
- CHENTSOV, N. N. **Statistical decision rules and optimal inference**. Estados Unidos da América: American Mathematical Society, 2000. v. 53. (Translations of Mathematical Monographs, v. 53). ISBN 3257227892. Citado na página 21.
- COVER, T. M.; THOMAS, J. A. **Elements of information theory**. 2. ed. Estados Unidos da América: Wiley-Interscience, 2006. ISBN 978-0-471-24195-9. Citado nas páginas 19, 20 e 43.
- DANI, S. G.; PAPADOPOULOS, A. **Geometry in History**. Suíça: Springer, 2019. ISBN 978-3030136086. Citado na página 48.
- DEGROOT, M. H.; SCHERVISH, M. J. **Probability and Statistics**. 4. ed. Estados Unidos da América: Addison Wesley Longman, 2011. ISBN 978-0321500465. Citado na página 41.
- DOUVEN, I. **The Art of Abduction**. Inglaterra: MIT Press, 2022. ISBN 9780262046701. Citado na página 35.
- DOWTY, J. G. Chentsov’s theorem for exponential families. p. 1–14, mar. 2018. Acesso em 21-03-2023. Disponível em: <<https://arxiv.org/abs/1701.08895>>. Citado na página 59.
- EISBERG, R.; RESNICK, R. **Física Quântica - Átomos, Moléculas, Sólidos, Núcleos e Partículas**. 1. ed. [S.l.]: Campus, 1979. ISBN 978-8570013095. Citado na página 33.
- ELSTON, R.; SATAGOPAN, J.; SUN, S. Genetic terminology. **Methods in molecular biology (Clifton, N.J.)**, v. 850, p. 1–9, dez. 2012. Acesso em 02-08-2022. Disponível em: <https://link.springer.com/protocol/10.1007/978-1-61779-555-8_1>. Citado na página 23.
- EUCLID. **Euclid’s elements of geometry**. 2008. Traduzido por Richard Fitzpatrick e editado por Johan L. Heiberg. ISBN 979-8589564587. Citado na página 48.
- FISHER, R. A. On the mathematical foundations of theoretical statistics. **Philosophical transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character**, The Royal Society London, v. 222, p. 309–368, jan. 1922. Acesso em 27-03-2023. Disponível em: <<https://royalsocietypublishing.org/doi/10.1098/rsta.1922.0009>>. Citado na página 66.
- FIUSA, G.; PINTO, D. O. S.; PIRES, D. P. Fidelity-based distance bounds for n-qubit approximate quantum error correction. **Physical Review A**, American Physical Society (APS), v. 107, n. 3, mar 2023. Disponível em: <<https://arxiv.org/abs/2212.04368>>. Citado na página 23.
- Ченцов, НН. Статические решающие правила и оптимальные выводы. Rússia: Наука, Физматлит, 1972. v. 4. (Теория вероятностей и математическая статистика, v. 4). Citado na página 21.

INSTITUTE, N. H. G. R. **Gamete**. 2023. Acesso em 02-08-2023. Disponível em: <<https://www.genome.gov/genetics-glossary/Gamete>>. Citado na página 23.

KLUNG, W.; CUMMINGS, M.; SPENCER, C.; PALLADINO, M.; KILLIAN, D. **Concepts of Genetics**. 12. ed. Estados Unidos da América: Pearson, 2019. ISBN 978-0134604718. Citado na página 23.

KOBAYOSHI, M. **Calibração de instrumentos de medição**. Brasil: SENAI-SP Editora, 2018. (Mecânica Dimensional). ISBN 9788565418157. Citado na página 38.

LAURITZEN, S. L. Statistical manifolds. In: **Differential Geometry in Statistical Inference**. Estados Unidos da América: Institute of Mathematical Statistics, 1987, (Institute of Mathematical Statistics, v. 10). cap. 4, p. 163–216. Citado nas páginas 22, 38, 44, 75 e 88.

LEWIS, F. P. History of the parallel postulate. **The American Mathematical Monthly**, Mathematical Association of America, v. 27, n. 1, p. 16–23, 1920. Acesso em 20-12-2022. Disponível em: <<https://www.jstor.org/stable/2973238?origin=crossref>>. Citado nas páginas 23 e 48.

LIU, J.; YUAN, H.; LU, X.-M.; WANG, X. Quantum fisher information matrix and multiparameter estimation. **Journal of Physics A: Mathematical and Theoretical**, IOP Publishing, v. 53, n. 2, p. 1–71, dez. 2019. Acesso em 31-05-2023. Disponível em: <<https://arxiv.org/abs/1907.08037>>. Citado na página 111.

MAGALHÃES, M. N. **Probabilidade e Variáveis Aleatórias**. 3. ed. Brasil: Edusp, 2011. ISBN 978-8531409455. Citado na página 36.

MAGNO, G. F. **Geometria da Informação: métricas em espaços de estado e correções na cota de Cramér-Rao**. Dissertação (Mestrado) — Instituto de Física de São Carlos, Universidade de São Paulo, Brasil, jul. 2020. Acesso em 19-12-2022. Citado nas páginas 20, 23, 25 e 38.

MAGRI, F. WDVV equations. **Il Nuovo Cimento C, SIF**, v. 38, n. 5, p. 1—10, abr. 2016. ISSN 03905551, 03905551. Acesso em 19-04-2023. Disponível em: <<https://www.sif.it/riviste/sif/ncc/econtents/2015/038/05/article/11>>. Citado nas páginas 103 e 104.

MANNING, C. D.; SCHÜTZE, H. **Foundations of Statistical Natural Language Processing**. Estados Unidos da América: MIT Press, 1999. ISBN 978-0262133609. Citado na página 43.

MOROZOVA, E. A.; CHENTSOV, N. N. Markov invariant geometry on manifolds of states. **Journal of Soviet Mathematics**, Springer, v. 56, p. 2648–2669, out. 1991. Acesso em 07-03-2023. Disponível em: <<https://link.springer.com/article/10.1007/BF01095975>>. Citado na página 86.

MURRAY, M. K.; RICE, J. W. **Differential geometry and statistics**. [S.l.]: Springer-Science+Business Media, B.V., 2017. v. 48. (Monographs on Statistics and Applied Probability, v. 48). Citado nas páginas 33 e 38.

NIELSEN, F. Legendre transformation and information geometry. **no. CIG-MEMO2**, p. 1–6, set. 2010. Acesso em 12-05-2023. Disponível em: <<https://doi.org/10.13140/RG.2.2.23048.67844>>. Citado nas páginas 99 e 100.

_____. **Geometric Structures of Information**. Suíça: Springer, 2018. (Signals and Communication Technology). ISBN 978-3030025199. Citado na página 35.

_____. An elementary introduction to information geometry. **Entropy**, MDPI AG, v. 22, n. 10, p. 1100, set. 2020. ISSN 1099-4300. Acesso em 10-03-2023. Disponível em: <<https://www.mdpi.com/1099-4300/22/10/1100>>. Citado nas páginas 28, 87 e 91.

_____. The many faces of information geometry. **Notices of the American Mathematical Society**, v. 69, p. 36–45, jan. 2022. Acesso em 15-06-2022. Disponível em: <<https://www.ams.org/notices/202201/rnoti-p36.pdf>>. Citado nas páginas 19, 20, 21, 22, 23, 43, 44, 46, 47, 67, 92, 94 e 95.

NIELSEN, M. A.; CHUANG, I. L. **Quantum Computation and Quantum Information**. Reino Unido: Cambridge University Press, 2010. ISBN 9781107002173. Citado na página 118.

OHARA, A. Doubly autoparallel structure on positive definite matrices and its applications. In: SPRINGER. **Geometric Science of Information: 4th International Conference, GSI 2019, Toulouse, France, August 27–29, 2019, Proceedings 4**. 2019. p. 251–260. ISBN 978-3-030-26980-7. Acesso em 25-04-2023. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-030-26980-7_26>. Citado nas páginas 96 e 97.

PETZ, D. Monotone metrics on matrix spaces. **Linear Algebra and its Applications**, v. 244, p. 81–96, 1996. Acesso em 27-05-2023. Disponível em: <https://www.sciencedirect.com/science/article/pii/0024379594002118?ref=pdf_download&fr=RR-2&rr=7cd7bf6d1d151b21>. Citado na página 110.

PICASSO, L. E. **Lectures in Quantum Mechanics: A Two-Term Course**. Suíça: Springer Cham, 2015. (UNITEXT for Physics). ISBN 9783319226323. Citado na página 107.

PIRES, D. P.; CIANCIARUSO, M.; CÉLERI, L. C.; ADESSO, G.; SOARES-PINTO, D. O. Generalized geometric quantum speed limits. **Physical Review X**, American Physical Society, v. 6, p. 021031, jun. 2016. Acesso em 18-05-2023. Disponível em: <<https://link.aps.org/doi/10.1103/PhysRevX.6.021031>>. Citado nas páginas 23, 29, 107, 108, 109, 110, 111, 112, 114, 115, 116, 117, 118, 119, 120, 121 e 122.

PISTONE, G.; SEMPI, C. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. **The Annals of Statistics**, Institute of Mathematical Statistics, v. 23, n. 5, p. 1543–1561, 1995. Acesso em 18-05-2023. Disponível em: <<https://projecteuclid.org/journals/annals-of-statistics/volume-23/issue-5/An-Infinite-Dimensional-Geometric-Structure-on-the-Space-of-all/10.1214/aos/1176324311.full>>. Citado na página 22.

PRICE, W. C.; CHISSICK, S. S.; HEISENBERG, W. **The Uncertainty Principle and Foundations of Quantum Mechanics: A Fifty Years' Survey**. Grã-Bretanha: Wiley-Interscience, 1977. ISBN 9780471994145. Citado na página 108.

RAO, C. R. Information and the accuracy attainable in the estimation of statistical parameters. In: CHAUDHURI, A.; GUPTA, S. N.; ROYCHOUDHURY, R. (Ed.). **A Tribute to the Legend of Professor C. R. Rao: The Centenary Volume**. Singapura: Springer Singapore, 2021. p. 1–13. ISBN 978-981-33-6991-7. Acesso em 02-08-2023. Disponível em: <https://link.springer.com/chapter/10.1007/978-981-33-6991-7_1>. Citado na página 20.

ROMAN, S. **Advanced Linear Algebra**. 3. ed. Estados Unidos da América: Springer, 2007. (Graduate Texts in Mathematics). ISBN 978-0387728285. Citado na página 53.

SCHWACHHÖFER, L.; AY, N.; JOST, J.; LÊ, H. V. Congruent families and invariant tensors. In: AY, N.; GIBILISCO, P.; MATÚŠ, F. (Ed.). **Information Geometry and Its Applications**. Suíça: Springer International Publishing, 2018. p. 157–187. ISBN 978-3-319-97798-0. Acesso em 25-03-2022. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-319-97798-0_6>. Citado nas páginas 64 e 65.

SHIMA, H.; YAGI, K. Geometry of hessian manifolds. **Differential Geometry and its Applications**, v. 7, p. 277–290, 1997. ISSN 0926-2245. Acesso em 03-05-2023. Disponível em: <<http://bit.ly/3Cu6VKt>>. Citado na página 95.

UOHASHI, K. α -connections and a symmetric cubic form on a riemannian manifold. **Entropy**, MDPI, v. 19, n. 7, p. 344, 2017. Acesso em 11-04-2023. Disponível em: <<https://www.mdpi.com/1099-4300/19/7/344>>. Citado na página 91.

WAN, K. K. **Quantum Mechanics: A Fundamental Approach**. Singapura: Jenny Stanford Publishing, 2019. ISBN 9781351333368. Citado na página 108.

WATERLOO, I. for Quantum Computing from University of. **Quantum 101 Glossary**. University of Waterloo, 2022. Acesso em 31-05-2023. Disponível em: <<https://uwaterloo.ca/institute-for-quantum-computing/quantum-101/quantum-101-glossary>>. Citado na página 118.

YEH, J. J. **Real Analysis: Theory of Measure and Integration**. 3. ed. Singapura: World Scientific Publishing Company, 2014. ISBN 3257227892. Citado na página 52.

ZHENG, S. **Fisher Information and Cramér-Rao Bound**. 2011. Acesso em 02-08-2023. Disponível em: <<https://shorturl.at/xMP34>>. Citado na página 38.

DEMONSTRAÇÃO DA CARACTERIZAÇÃO PARA O TENSOR DE AMARI-CHENTSOV EM $\mathcal{P}_+(\mathcal{X})$

Teorema 7. (AY *et al.*, 2017) Atribua a cada conjunto finito não vazio \mathcal{X} um campo tensorial 3-covariante não trivial $S^{\mathcal{X}}$ em $\mathcal{P}_+(\mathcal{X})$. Se, para cada núcleo de Markov congruente $K: \mathcal{X} \rightarrow \mathcal{X}'$, houver a invariância no sentido

$$S_{\xi}^{\mathcal{X}}(A, B, C) = S_{K_*(\xi)}^{\mathcal{X}'}(d_{\xi}K_*(A), d_{\xi}K_*(B), d_{\xi}K_*(C)), \quad (\text{A.1})$$

então existe uma constante $\alpha \neq 0$ tal que $S^{\mathcal{X}} = \alpha T^{\mathcal{X}}$ para todo \mathcal{X} , onde $T^{\mathcal{X}}$ denota o tensor de Amari-Chentsov em $\mathcal{P}_+(\mathcal{X})$.

Demonstração. Sigamos passos análogos aqueles realizados na demonstração de 2. Seja, então, $\sigma: \mathcal{X} \rightarrow \mathcal{X}$ uma permutação em \mathcal{X} e defina

$$c_{\mathcal{X}} := \frac{1}{n} \sum_{i \in I} \delta^i, \quad n := |\mathcal{X}|.$$

Assim, de 2, sabemos que $c_{\mathcal{X}}$ é invariante à esquerda por σ e que, dado $E_i := (c_{\mathcal{X}}, \delta^i - c_{\mathcal{X}}) \in T_{c_{\mathcal{X}}} \mathcal{P}_+(\mathcal{X})$, com $i \in J$, então

$$d_{c_{\mathcal{X}}} \sigma_*(E_i) = E_{\sigma(i)}.$$

Dito isso, para cada $i, j \in J$, com $i \neq j$, tome σ a transposição de x_i e x_j . Calculando, então, $S_{iii}^{\mathcal{X}}(c_{\mathcal{X}})$ sob a hipótese da invariância A.1, obtemos

$$\begin{aligned} S_{iii}^{\mathcal{X}}(c_{\mathcal{X}}) &= S_{c_{\mathcal{X}}}^{\mathcal{X}}(E_i, E_i, E_i) \\ &= S_{c_{\mathcal{X}}}^{\mathcal{X}}(d_{c_{\mathcal{X}}} \sigma_*(E_j), d_{c_{\mathcal{X}}} \sigma_*(E_i), d_{c_{\mathcal{X}}} \sigma_*(E_i)) \\ &= S_{c_{\mathcal{X}}}^{\mathcal{X}}(E_{\sigma(i)}, E_{\sigma(i)}, E_{\sigma(i)}) \\ &= S_{c_{\mathcal{X}}}^{\mathcal{X}}(E_j, E_j, E_j) \\ &= S_{jjj}^{\mathcal{X}}(c_{\mathcal{X}}) =: \tilde{f}_1(n). \end{aligned}$$

De forma análoga, calculando $S_{iik}^{\mathcal{X}}(c_{\mathcal{X}})$ com $i \neq k$, temos que

$$\begin{aligned} S_{iik}^{\mathcal{X}}(c_{\mathcal{X}}) &= S_{c_{\mathcal{X}}}^{\mathcal{X}}(\mathbf{d}_{c_{\mathcal{X}}}\sigma_{\mathcal{X}}(E_i), \mathbf{d}_{c_{\mathcal{X}}}\sigma_{\mathcal{X}}(E_i), \mathbf{d}_{c_{\mathcal{X}}}\sigma_{\mathcal{X}}(E_k)) \\ &= S_{c_{\mathcal{X}}}^{\mathcal{X}}(E_j, E_j, E_k) \\ &= S_{jjk}^{\mathcal{X}}(c_{\mathcal{X}}) =: \tilde{f}_2(n). \end{aligned}$$

Além disso, para $i \neq j \neq k \neq i$, observe que os termos $S_{ijk}^{\mathcal{X}}(c_{\mathcal{X}}) =: \tilde{f}_3(n)$ do tensor $S_{c_{\mathcal{X}}}^{\mathcal{X}}$ coincidem. Note que, por um lado,

$$\sum_{j,k \in J} S_{ijk}^{\mathcal{X}}(c_{\mathcal{X}}) = \tilde{f}_1(n) + 3(n-1)\tilde{f}_2(n) + (n-1)(n-2)\tilde{f}_3(n),$$

enquanto que, por outro,

$$\begin{aligned} \sum_{j,k \in J} S_{ijk}^{\mathcal{X}}(c_{\mathcal{X}}) &= \sum_{j,k \in J} S_{c_{\mathcal{X}}}^{\mathcal{X}}(E_i, E_j, E_k) \\ &= \sum_{j \in J} S_{c_{\mathcal{X}}}^{\mathcal{X}}\left(E_i, E_j, \sum_{k \in J} E_k\right) \\ &= 0. \end{aligned}$$

Assim, obtemos que

$$\tilde{f}_1(n) + 3(n-1)\tilde{f}_2(n) + (n-1)(n-2)\tilde{f}_3(n) = 0. \quad (\text{A.2})$$

Analogamente,

$$0 = \sum_{j \in J} S_{iij}^{\mathcal{X}}(c_{\mathcal{X}}) = \tilde{f}_1(n) + (n-1)\tilde{f}_2(n). \quad (\text{A.3})$$

Dessa maneira, usando A.2 e A.3, podemos escrever $\tilde{f}_1(n)$ e $\tilde{f}_2(n)$ em termos de $\tilde{f}_3(n)$, respectivamente, como

$$\tilde{f}_1(n) = \frac{(n^2 - 3n + 2)}{2}\tilde{f}_3(n), \quad \tilde{f}_2(n) = \frac{(2-n)}{2}\tilde{f}_3(n). \quad (\text{A.4})$$

Posto isso, temos, para todo

$$A = \left(c_{\mathcal{X}}, \sum_{i \in J} a_i \delta^i\right), B = \left(c_{\mathcal{X}}, \sum_{i \in J} b_i \delta^i\right), C = \left(c_{\mathcal{X}}, \sum_{i \in J} c_i \delta^i\right) \in T_{c_{\mathcal{X}}}\mathcal{P}_+(\mathcal{X}),$$

as igualdades

$$\begin{aligned} S_{c_{\mathcal{X}}}^{\mathcal{X}}(A, B, C) &= \sum_{i,j,k \in J} S_{ijk}^{\mathcal{X}}(c_{\mathcal{X}}) a_i b_j c_k \\ &= \tilde{f}_1(n) \sum_{i \in J} a_i b_i c_i + \frac{(2-n)}{2} \tilde{f}_3(n) \sum_{\substack{i,j \in J \\ i \neq j}} (a_i b_i c_j + a_i b_j c_i + a_j b_i c_i) + \tilde{f}_3(n) \sum_{\substack{i,j,k \in J \\ i \neq j \neq k \neq i}} a_i b_j c_k, \end{aligned}$$

ou seja,

$$S_{c_{\mathcal{X}}}^{\mathcal{X}}(A, B, C) = \tilde{f}_3(n) \left(\frac{(n^2 - 3n + 2)}{2} \sum_{i \in J} a_i b_i c_i + \tilde{f}_2(n) \sum_{\substack{i,j \in J \\ i \neq j}} (a_i b_i c_j + a_i b_j c_i + a_j b_i c_i) + \sum_{\substack{i,j,k \in J \\ i \neq j \neq k \neq i}} a_i b_j c_k \right), \quad (\text{A.5})$$

onde o último termo dentro do parênteses computa os termos de $S_{c\mathcal{X}}^{\mathcal{X}}$ com todos os seus índices distintos. Perceba que

$$\begin{aligned} -\frac{3n}{2} \sum_{i \in J} a_i b_i c_i - \frac{n}{2} \sum_{\substack{i, j \in J \\ i \neq j}} (a_i b_i c_j + a_i b_j c_i + a_j b_i c_i) &= -\frac{n}{2} \sum_{i, j \in J} (a_i b_i c_j + a_i b_j c_i + a_j b_i c_i) \\ &= -\frac{n}{2} \sum_{i \in J} \left(a_i b_i \sum_{j \in J} c_j + a_i c_i \sum_{j \in J} b_j + b_i c_i \sum_{j \in J} a_j \right) \\ &= 0. \end{aligned}$$

Além disso,

$$\begin{aligned} \sum_{i \in J} a_i b_i c_i + \sum_{\substack{i, j \in J \\ i \neq j}} (a_i b_i c_j + a_i b_j c_i + a_j b_i c_i) + \sum_{\substack{i, j, k \in J \\ i \neq j \neq k \neq i}} a_i b_j c_k &= \sum_{i, j, k \in J} a_i b_j c_k \\ &= \sum_{i, j \in J} a_i b_j \sum_{k \in J} c_k \\ &= 0. \end{aligned}$$

Então, A.5 pode ser reescrita como

$$\begin{aligned} S_{c\mathcal{X}}^{\mathcal{X}}(A, B, C) &= \frac{n^2}{2} \tilde{f}_3(n) \sum_{i \in J} a_i b_i c_i \\ &= \frac{\tilde{f}_3(n)}{2} \sum_{i \in J} \frac{a_i b_i c_i}{\frac{1}{n^2}} \\ &= \frac{\tilde{f}_3(n)}{2} T_{c\mathcal{X}}^{\mathcal{X}}(A, B, C). \end{aligned}$$

Vamos, agora, mostrar que $\tilde{f}_3(n)$ é constante. A partir deste ponto, os argumentos da demonstração seguem quase que totalmente iguais aos da demonstração do teorema 2. Destacado isso, seja $\tilde{\mathcal{X}} = \{\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_k\}$ conjunto finito não vazio e defina $\mathcal{X}' := \mathcal{X} \times \tilde{\mathcal{X}}$. Considere a partição de \mathcal{X}' cujos elementos são conjuntos da forma $A_i := \{(x_i, \tilde{x}_j) \in \mathcal{X}' \mid 1 \leq j \leq k\}$. Desse modo, definimos o núcleo de Markov

$$K : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{X}'), \quad x_i \mapsto K^i = \frac{1}{k} \sum_{j=1}^k \delta^{(i,j)},$$

cujo morfismo induzido é dado por

$$K_*(c_{\mathcal{X}}) = \frac{1}{nk} \sum_{i \in J} \sum_{j=1}^k \delta^{(i,j)} = c_{\mathcal{X}'}$$

Defina, também, $E'_{(i,j)} := (c_{\mathcal{X}'}, \delta^{(i,j)} - c_{\mathcal{X}'}) \in T_{c_{\mathcal{X}'}} \mathcal{P}_+(\mathcal{X}')$. Então, conforma calculado na demonstração de 2,

$$d_{c_{\mathcal{X}}} K_*(E_i) = \frac{1}{k} \sum_{j=1}^k E'_{(i,j)}.$$

Assim, sendo $q, r, s \in J$ com $q \neq r \neq s \neq q$, temos que

$$S_{c_{\mathcal{X}}}^{\mathcal{X}}(E_q, E_r, E_s) = \frac{\tilde{f}_3(n)}{2} T_{c_{\mathcal{X}}}^{\mathcal{X}}(E_q, E_r, E_s) = \tilde{f}_3(n).$$

Consequentemente, utilizando a invariância A.1, obtemos

$$\begin{aligned} \tilde{f}_3(n) &= S_{c_{\mathcal{X}'}}^{\mathcal{X}'} \left(\frac{1}{k} \sum_{l=1}^k E'_{(q,l)}, \frac{1}{k} \sum_{r=1}^k E'_{(r,m)}, \frac{1}{k} \sum_{p=1}^k E'_{(s,p)} \right) \\ &= \frac{1}{k^3} \sum_{l,m,p=1}^k S_{c_{\mathcal{X}'}}^{\mathcal{X}'} \left(E'_{(q,l)}, E'_{(r,m)}, E'_{(s,p)} \right) \\ &= \tilde{f}_3(nk). \end{aligned}$$

Invertendo a função de n e k , obtemos $\tilde{f}_3(k) = \tilde{f}_3(nk) = \tilde{f}_3(n)$, de forma que fica mostrado que \tilde{f}_3 é independente da cardinalidade de \mathcal{X} . Além disso, da não trivialidade de $S^{\mathcal{X}}$, temos que $\frac{\tilde{f}_3(n)}{2} \neq 0$. Denotando $\frac{\tilde{f}_3(n)}{2}$ por α , vamos estender a igualdade $S^{\mathcal{X}} = \alpha T^{\mathcal{X}}$ para todos os pontos de $\mathcal{P}_+(\mathcal{X})$. Para tal, considere $\xi \in \mathcal{P}_+(\mathcal{X})$ cujas coordenadas na base canônica de $\mathcal{S}(\mathcal{X})$ são racionais, isto é,

$$\xi = \sum_{i \in J} \frac{k_i}{n} \delta^i, \quad \sum_{i \in J} k_i = n.$$

Sejam, também, $\widetilde{\mathcal{X}}_i$ conjuntos de cardinalidade k_i , com $i \in J$, e que denotaremos por $\widetilde{\mathcal{X}}_i := \{\widetilde{x}_1^i, \widetilde{x}_2^i, \dots, \widetilde{x}_{k_i}^i\}$. Desse modo, defina o conjunto

$$\mathcal{X}' := \bigcup_{i=1}^n \left(\{x_i\} \times \widetilde{\mathcal{X}}_i \right)$$

e considere o núcleo da Markov

$$K : \mathcal{X} \rightarrow \mathcal{P}_+(\mathcal{X}'), \quad x_i \mapsto \frac{1}{k_i} \sum_{j=1}^{k_i} \delta^{(i,j)}. \quad (\text{A.6})$$

Por 2, sabemos que

$$K_*(\xi) = c_{\mathcal{X}'}$$

Assim, dado $A = (\xi, \sum_{i \in J} a_i \delta^i) \in \mathcal{P}_+(\mathcal{X})$, temos que

$$d_{\xi} K_*(A) = (c_{\mathcal{X}'}, K_*(a)) = \left(c_{\mathcal{X}'}, \sum_{i \in J} \sum_{j=1}^{k_i} \frac{a_i}{k_i} \delta^{(i,j)} \right).$$

Aplicando mais uma vez a invariância 3.6, obtemos, para todo

$$A = \left(\xi, \sum_{i \in J} a_i \delta^i \right), B = \left(\xi, \sum_{i \in J} b_i \delta^i \right), C = \left(\xi, \sum_{i \in J} c_i \delta^i \right) \in \mathcal{P}_+(\mathcal{X}),$$

que

$$\begin{aligned} S_{\xi}^{\mathcal{X}}(A, B, C) &= S_{K_*(\xi)}^{\mathcal{X}'}(d_{\xi} K_*(A), d_{\xi} K_*(B), d_{\xi} K_*(C)) \\ &= \alpha T_{c_{\mathcal{X}'}}^{\mathcal{X}'}(d_{\xi} K_*(A), d_{\xi} K_*(B), d_{\xi} K_*(C)) \\ &= \alpha \sum_{i \in J} n^2 \frac{a_i}{k_i} \frac{b_i}{k_i} \frac{c_i}{k_i} k_i \\ &= \alpha T_{\xi}^{\mathcal{X}}(A, B, C). \end{aligned}$$

Devido ao fato de [A.6](#) sempre poder ser construído para todo $\xi \in \mathcal{P}_+(\mathcal{X})$ com coordenadas racionais, temos que a sequência de igualdades acima é verdadeira para todo ξ nessas condições. Como $S^{\mathcal{X}}$ e $\alpha T^{\mathcal{X}}$ coincidem nesses pontos e como \mathbb{Q} é denso em \mathbb{R} , temos que $S^{\mathcal{X}} = \alpha T^{\mathcal{X}}$. \square

