**UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS**

Gabriel Fukamoto Magno

Geometria da Informação:
métricas em espaços de estado e correções na cota de Cramér-Rao

São Carlos

2019

Gabriel Fukamoto Magno

Geometria da Informação:
métricas em espaços de estado e correções na cota de Cramér-Rao

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de concentração: Física Teórica e Experimental

Orientador: Prof. Dr. Diogo de Oliveira Soares Pinto

Coorientador: Prof. Dr. Carlos Henrique Grossi Ferreira

Versão corrigida
(versão original disponível na Unidade que aloja o Programa)

São Carlos
2019

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Magno, Gabriel FuKamoto

Geometria da Informação: métricas em espaços de estado e correções na cota de Cramér-Rao / Gabriel FuKamoto Magno; orientador Diogo de Oliveira Soares Pinto; co-orientador Carlos Henrique Grossi Ferreira -- São Carlos, 2019.
158 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Física Teórica e Experimental) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2019.

1. Geometria. 2. Teoria de Informação. 3. Cota de Cramér-Rao. I. Pinto, Diogo de Oliveira Soares, orient. II. Ferreira, Carlos Henrique Grossi, co-orient. III. Título.

*Aos meus avós,
Mituko e Hitoxi*

AGRADECIMENTOS

À Universidade de São Paulo. Em defesa da autonomia da universidade pública, gratuita e de qualidade.

Ao Prof. Dr. Diogo de Oliveira Soares Pinto pela orientação, pela liberdade e estímulo que me proporcionou para trabalhar com a matemática que tanto me agrada, por ter me acalmado em momentos de ansiedade, pelo suporte acadêmico que sempre esteve disposto a me fornecer e por ser um cara humano e companheiro mesmo dentro deste mundo produtivista.

Ao Prof. Dr. Eduardo Azevedo que disponibilizou o espaço físico onde pudesse trabalhar e que certas vezes aparecia para dar uma de suas risadas características.

Ao Prof. Dr. Carlos Grossi Ferreira e o amigo Matheus Costa que auxiliaram com o formalismo matemático. Ao Prof. Dr. Daniel Vanzella pelo auxílio com contas técnicas.

O meu respeito a todos aquele que contribuíram com a minha formação intelectual e cidadã, com destaque ao careca Prof. Fernando Pucinelli.

Ao CEFER USP - São Carlos, Hevert e Gizele pelos treinamentos físicos. À Kim pelos treinos de boxe, esporte que é minha paixão. Afinal, atividade física oxigena o cérebro e ajuda a pensar.

Aos meus pais, Márcia e José Aguinaldo, e minha irmã Mariana que acompanharam toda minha caminhada até aqui. Especialmente aos meus pais, que sempre foram meu esteio, apoiaram as decisões que tomei acerca da vida, foram formadores de muitos princípios e permanecem ao meu lado auxiliando e desejando sempre o melhor.

À Amanda, minha senhora, que foi parte integral de minha vida todos esses anos de graduação e pós. Obrigado por ter permanecido ao meu lado em tempos bons, nem tão bons assim, fáceis e difíceis. Pelas conversas magníficas sobre os assuntos mais diversos possíveis, com destaque aqueles sobre existencialismo e políticas sociais, e por me tornar uma pessoa melhor.

O principal agradecimento fica ao meus avós *Mituko e Hitoxi*, respectivamente, professora da rede pública e encanador/eletricista aposentados. Pessoas que tiveram vida simples e, a custo de certo conforto, proporcionaram a melhor educação que podiam para seu neto que hoje humildemente tenta retribuir tornando-se um mestre em ciências. Qualquer conquista na minha vida será de vocês dois também, obrigado por confiar em mim, vou levar vocês sempre em minha vida, sou exorbitantemente grato.

Esta dissertação foi financiada pelo CAPES.

“Matemática pura e física estão se tornando cada vez mais conectadas, apesar das suas metodologias continuarem diferentes. Pode-se descrever a situação dizendo que matemáticos jogam um jogo em que eles mesmos inventam as regras, enquanto os físicos jogam um jogo em que as regras são ditadas pela Natureza, mas conforme o tempo passa torna-se cada vez mais evidente que as regras que os matemáticos acham interessantes são as mesmas que a Natureza escolheu.”

P. A. M. Dirac

(traduzido pelo autor)

RESUMO

MAGNO, G. F. **Geometria da Informação: métricas em espaços de estado e correção na cota de Cramér-Rao**. 2019. 154p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2019.

A Geometria da informação promove uma investigação da estrutura geométrica de variedades estatísticas, fornecendo uma série de elucidações em áreas que vão desde ciências da informação até ciências físicas, além de resultados aplicáveis a problemas como os de inferência estatística. Neste texto, vamos apresentar a estrutura dualística (métrica de Fisher, $\pm\alpha$ -conexões) na variedade de probabilidades clássicas vinda de f -divergências globalmente definidas por funções convexas. Vamos então estender este formalismo para produzir o seu análogo quântico, obtendo uma família de métricas de Fisher não comutativas. Em seguida, relacionamos estas métricas com as métricas monotônicas permitidas na mecânica quântica pelo do Teorema de Morozova-Chentsov-Petz via conexão operadores função convexa e monotônica. Por fim, mostramos uma aplicação das técnicas geométricas em um problema de inferência estatística, onde calcularemos correções de ordem superior na cota de Cramér-Rao para um estimador de um parâmetro a partir da estrutura geométrica do espaço de Hilbert adjacente, clássico ou quântico, observando as intersecções deste processo com os resultados anteriormente apresentados.

Palavras-chave: Geometria. Teoria de informação. Cota de Cramér-Rao.

ABSTRACT

MAGNO, G. F. **Information Geometry: metrics in state spaces and corrections to the Cramér-Rao bound.** 2019. 154p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2019.

Information geometry promotes an investigation of the geometric structure of statistical manifolds, providing a series of elucidations in areas ranging from information science to physical sciences. In this text, we show a dualistic structure (Fisher metric, $\pm\alpha$ -connections) on the classical probability manifold coming from f -divergences globally defined by convex functions. Thus, we extend this formalism to obtain its quantum analog, finding a noncommutative family of Fisher metrics. Then, we connect these metrics with the allowed monotone metrics in quantum mechanics from the Morozova-Chentsov-Petz theorem by a relation between operator convex and monotone functions. At last, we show one application of geometric technics on a statistical inference problem, where we are going to calculate higher order correction on the Cramér-Rao bound for an uniparametric estimator from the geometric structure of the adjacent Hilbert space, classical or quantum, observing the intersections with the results previously presented.

Keywords:Geometry. Information theory. Cramér-Rao bound

LISTA DE ILUSTRAÇÕES

Figura 1 – Carta $(\mathcal{O}, \phi, x^i)$ no ponto p da variedade topológica M	24
Figura 2 – Funções de transição entre cartas compatíveis.	25
Figura 3 – Conexão afim vinda de uma abordagem usando infinitesimais.	36
Figura 4 – $\mathcal{P} \in \mathcal{H}$ folheado por superfícies a τ constante. Para uma família uniparamétrica qualquer $\dot{\xi}^a$ e $\nabla^a \tau$ não são paralelos; apenas famílias exponenciais, geradas a partir de um ponto q , cruzam perpendicularmente as superfícies τ	80

SUMÁRIO

1	INTRODUÇÃO	19
2	TÓPICOS ELEMENTARES EM GEOMETRIA DIFERENCIAL	23
2.1	Variedades diferenciáveis	23
2.2	Curvas e mapas diferenciáveis	25
2.2.1	Funções suaves e difeomorfismos	25
2.2.2	Curva parametrizada suave	26
2.3	Espaço tangente, cotangente e tensores	26
2.3.1	Vetores Tangentes	26
2.3.2	Espaço cotangente e tensores	28
2.3.3	Campos vetoriais e tensoriais	30
2.3.4	Transformação de coordenadas	31
2.3.5	Comentário sobre fibrados	32
2.4	Conexões e breve introdução a geometria Riemanniana	33
2.4.1	Variedade pseudo-Riemanniana e Riemanniana	33
2.4.2	Conexões afins e derivada covariante	34
2.4.3	Planicidade, torção e curvatura	38
2.4.4	Conexão Riemanniana	39
3	GEOMETRIA DE MODELOS ESTATÍSTICOS CLÁSSICOS E QUÂN- TICOS	43
3.1	Modelos estatísticos	43
3.2	Geometria de modelos clássicos, métrica de Fisher e α-conexões	45
3.2.1	Matriz informação de Fisher	45
3.2.2	Estatística suficiente	45
3.2.3	As α -conexões	46
3.3	Discussão sobre a origem afim das α-conexões	47
3.4	Conexões Duais	48
3.5	α-divergências em modelos estatísticos	49
3.5.1	Divergências	49
3.5.2	As α -divergências	52
3.6	Geometria da informação para sistemas quânticos	55
3.6.1	Espaço de estados	55
3.6.2	Caracterização do espaço tangente	57
3.6.3	Divergência e métricas quânticas	60
3.6.4	Discussão sobre a contratividade das α -divergências	64

4	O TEOREMA DE MCP E A RELAÇÃO ENTRE OPERADOR FUNÇÃO MONOTÔNICA E CONVEXA	65
4.1	O Teorema de MCP	65
4.2	MCP, dinâmica unitária e Wigner-Yanase	68
4.2.1	Coeficientes da métrica	68
4.2.2	Métrica de Wigner-Yanase na dinâmica unitária	69
4.3	Relação operador função monotônica e convexa	71
5	CORREÇÕES NA COTA DE CRAMÉR-RAO	75
5.1	Estatística clássica de um parâmetro	75
5.1.1	Estatística usual na notação indexial	75
5.1.2	Métrica de Fisher	76
5.1.3	Cota de Cramér-Rao e famílias exponenciais	78
5.1.4	Cotas de Bhattacharyya	82
5.2	Estatística quântica de um parâmetro, caso puro	84
5.2.1	Geometria de estado puro	84
5.2.2	Dinâmica do tipo Schrödinger	87
5.2.3	Estimativa quântica	90
5.2.4	Correções de ordem superior na cota da variância quântica	92
5.3	A correção no caso misto	98
5.3.1	Adaptação da notação	98
5.3.2	Estimativa quântica no caso mistos	100
5.3.3	Correções de ordem superior na cota da variância para o caso misto	102
6	CONCLUSÃO	109
	REFERÊNCIAS	111
	APÊNDICES	117
	APÊNDICE A – ELEMENTOS DE TOPOLOGIA GERAL E TEN- SORES	119
A.1	Topologia	119
A.1.1	Espaços topológicos	119
A.1.2	Fecho, vizinhança e convergência	120
A.1.3	Homeomorfismos	121
A.1.4	Espaço de Hausdorff	122
A.2	Tensores	122

A.2.1	Bases, componentes e mudança de base - componentes contravariantes e covariantes	122
A.2.2	Tensor	125
A.2.3	Tensores simétricos e antissimétricos em $T^{(k,0)}(V)$ e $T^{(0,k)}(V)$	126
A.2.4	Sobe e desce de índices	127
	APÊNDICE B – METROLOGIA	131
B.1	Metrologia clássica	131
B.1.1	Contextualização	131
B.1.2	Informação de Fisher	133
B.1.2.1	Caso de uma única observação	133
B.1.2.2	Transformação de parâmetro	134
B.1.2.3	Caso de n observações	134
B.1.3	A cota de Cramér-Rao	135
B.1.4	O teorema da distribuição assintótica do MLE	136
B.2	Metrologia quântica de estado puro	137
B.2.1	Contextualização	137
B.2.2	Cota de Cramér-Rao quântica	138
B.2.3	Informação de Fisher para estados puros e evoluções unitárias	138
B.2.4	Emaranhamento e o CRB	140
	APÊNDICE C – COTA DE CRAMÉR-RAO DO PONTO DE VISTA GEOMÉTRICO	143
C.1	Famílias exponencial e mistura	143
C.2	Caso da variedade clássica	144
C.3	Caso quântico	146
C.3.1	Covariâncias generalizadas	146
C.3.2	As cotas quânticas	148
	APÊNDICE D – MÉTRICA DE FUBINI-STUDY	149
D.1	Espaço projetivo	149
D.2	Fubini-study no espaço de estados quânticos puros uniparamétricos	153

1 INTRODUÇÃO

Falando um pouco de geometria, um bom exercício é pensar que pilares da física como relatividade, mecânica quântica, teoria de campos, entre outros, têm suas estruturas matemáticas baseadas em teoremas, digamos, excêntricos formulados por um conjunto de matemáticos mais excêntricos ainda.⁴⁻⁶ Anos de questionamentos rigorosos acerca do quinto postulado de Euclides,⁷ que por sua vez aparentava certa obviedade, levaram a inauguração de novas geometrias que foram cruciais nas teorias físicas de Einstein e dos espaços de Hilbert.

Assim, podemos pensar que estruturas geométricas na Física são a grande “arena” na qual descrevemos a evolução dos sistemas,¹ portanto, é de grande sensatez o estudo do formalismo matemático-geométrico presente na descrição dos fenômenos físicos. A destacar o fato de que estruturas geométricas se manifestarem desde o estudo de cálculo básico, até na geometria do espaço de Minkowski,⁸⁻¹¹ estrutura simplética da mecânica clássica¹² ou na estrutura de produto interno do espaço de Hilbert.¹³

Agora, falando de ciência da informação quântica, vemos que ela é um campo interdisciplinar moderno de pesquisa nas áreas de física, matemática, engenharia e computação.^{14,15} Sua proposta passa por estudar e esclarecer as leis da mecânica quântica podendo idealizar protocolos e critérios que maximizem a eficiência, aquisição e processamento de informação. Sob a luz da teoria de informação, o desenvolvimento destes aspectos permite a proposição de novas ideias e materiais técnicos com caráter robusto e universal para ser aplicado à implementação física. Neste ponto, vale ressaltar seu grande potencial de pesquisa experimental e aplicações práticas,¹⁶⁻²² visando produzir tecnologia através de montagens que implementem os protocolos de manipulação dos estados quânticos com possibilidade de impacto na vida socioeconômica presente e futura.

Acerca da abordagem conceitual da área, vamos nos concentrar na discussão sobre medidas de distinguibilidade em espaços de estado, que proporcionam um paralelo direto com as questões geométricas. Dado um sistema quântico – uma possível tecnologia quântica – temos que os agentes fundamentais responsáveis pelo processamento da informação no sistema são os estados quânticos. Conhecidos os processos que agem sobre os estados, manipular adequadamente estes processos descritos por operações quânticas permite uma otimização na transmissão de dados. Afinal, a ação de qualquer protocolo deve codificar informação no estado do sistema. Ter acesso a esta informação codificada, após a aplicação da operação, pede por alguma maneira de distinguir os estados inicial e resultante do sistema.²³

Distinguir estados em modelos probabilísticos quânticos se mostra uma tarefa

típica na teoria da informação. São conhecidas diversas medidas de distância em espaços de estados na literatura que são utilizadas como quantificadores de distinguibilidade, por exemplo, a distância do traço, distância de Hellinger, distância de Hilbert-Schmidt, entropias relativas, entre outras.^{24,25} Como o conceito de distância é intimamente ligado a algum tipo de estrutura geométrica do espaço envolvido, temos o rastro da conexão entre entre métodos de geometria e teoria de informação.

Historicamente, na década 80, Wootters²⁶ começa os primeiros estudos acerca do elo entre distâncias estatísticas em espaços de probabilidade com medidas de distinguibilidade. Alguns anos mais tarde, o problema foi trazido para a roupagem de geometria diferencial – citando alguns: Morozava & Chentsov,²⁷ Amari,² Braunstein & Caves²⁸ e Petz²⁹ – caracterizando distinguibilidade estatística por meio de métricas Riemannianas em espaços clássicos e quânticos.

Esta formulação geométrica da informação vai tratar então tanto o espaço de estados clássico – conjunto formado por funções densidade de probabilidade – quanto quântico – conjunto de operadores densidade que atuam sobre o espaço de Hilbert do sistema – como variedades diferenciáveis equipadas com tensores métricos que irão definir a noção de distância entre os elementos da variedade. Tal abordagem permitirá o desenvolvimento de resultados de interesse físico explorando propriedades métricas dos espaço nos levando a um novo grau de profundidade na visão que temos dos mesmos.²

Desta forma, fica natural a interpretação de distinguibilidade como métricas definidas em variedades estatísticas. Isso culmina na geometria da informação, área que aplica métodos de geometria diferencial na solução e formalização de problemas da ciência da informação, obtendo resultados robustos e aplicáveis a questões de interesse, como por exemplo inferência estatística.

Esta dissertação inicialmente irá se dedicar a apresentar resultados já conhecidos sobre variedades estatísticas clássicas e quânticas, examinando e relacionando suas estruturas métricas da maneira mais sistemática possível. Desta maneira podemos ver como esses resultados vão aparecendo quando aplicarmos o formalismo geométrico a um problema de interesse em inferência estatística, onde queremos calcular correções de ordem superior para a cota de Cramér-Rao no espaço de estados quânticos mistos, usando passos semelhantes aos que já foram feitos no cálculo das correções clássicas e quânticas de estado puro. Vamos ver durante o processo que conforme vão surgindo intersecções entre os resultados, conclusões físicas podem ser tiradas com embasamento geométrico.

Vale notar que em geral nossas discussões serão válidas em espaços de dimensão finita.* No caso quântico exigiremos, quando necessário, espectros não degenerados por

*No caso de espaços clássicos significa que é interessante trabalharmos com amostragens finitas, já em espaços quânticos trabalhamos com operadores pertencentes ao conjunto de matrizes $n \times n$ com entradas complexas, denotado por $M_n(\mathbb{C})$.

simplicidade. Apontaremos mais à frente que dimensões infinitas são perigosas e não nos fornecem estrutura de variedade. Além disso, nossa discussão está situada no regime quântico não-relativístico tradicional.

O texto contém resultados que foram trabalhados durante aproximadamente 18 meses no projeto de pesquisa relativo a geometria da informação e esta organizado como segue:

No capítulo 2 apresentamos os principais conceitos matemáticos que estarão presentes durante toda a discussão, começando pelas definições básicas acerca de variedades diferenciáveis e espaços tangentes até chegar em conexões e uma breve introdução a geometria Riemanniana.

No capítulo 3 iniciamos a apresentação dos estudos desenvolvidos sobre variedades estatísticas. Começamos definindo um modelo estatístico e sua estrutura de variedade suave. Trabalharemos com o conjunto probabilidades clássicas que possuem uma métrica privilegiada na sua variedade pelo Teorema de Chentsov, a métrica de Fisher. Apresentaremos as α -conexões desta variedade e sua origem afim, mostrando que junto da métrica de Fisher elas formam uma estrutura dualística livre de torção que, por sua vez, pode ser obtida através de uma α -divergência globalmente definida através de funções convexas, estas divergências ainda terão formatos já vistos na literatura geral. Em seguida, queremos aplicar o que aprendemos no caso clássico a variedade formada pelo conjunto operadores densidade da Mecânica Quântica introduzindo um análogo quântico da estrutura dualística Fisher e α -conexões. Definindo uma α -divergência quântica veremos que ela de fato fornece um análogo ao caso clássico valendo também a origem afim das conexões, de onde obteremos uma família de métricas para direções não-comutativas do espaço de estados e para a direções comutativas recuperaremos o formato da informação de Fisher clássica.

No capítulo 4 vamos estabelecer a ligação entre as métricas quânticas obtidas no capítulo anterior com as métricas contrativas permitidas pelo Teorema de MCP. Esta ligação será feita através de uma correspondência entre operadores função convexa e operadores função monotônica que definem as métricas permitidas quântica, mostrando que o conjunto de métricas de Fisher não-comutativas, satisfazendo algumas hipóteses, deve cair dentro das métricas previstas por MCP.

No capítulo 5 abordaremos um problema de inferência estatística, que consiste no cálculo de correções na cota da variância de um estimador uniparamétrico não viesado, utilizando da estrutura geométrica dos espaços envolvidos. Veremos que intersecções com os resultados discutidos nos dois capítulos anteriores aparecerão de forma natural. Começaremos abordando o espaço de probabilidades clássicas de onde seguirão as correções do tipo Bhattacharyya. Introduziremos então uma estrutura complexa no espaço clássico recuperando dinâmica do tipo Schrödinger e permitindo o cálculo de correções para o caso de estados quânticos puro. Por fim, apresentaremos a adaptação do processo ao caso misto,

vendo que a correção obtida é muito parecida com o caso puro mas não se reduz a ele trivialmente.

No capítulo 6 temos as conclusões do trabalho, apresentando sugestões possíveis para trabalhos futuros a partir dos nossos resultados.

Os apêndices A, B, C, D, apesar de se terem o rótulo de apêndices, são complementares à leitura. Contém resultados que enriquecem as interpretações desenvolvidas ao longo dos capítulos, além de detalhes técnicos omitidos no texto. Recomendamos fortemente a leitura.

Observação 1. Se o leitor quiser acusar este trabalho de ser apenas um exercício de leitura do autor, este não irá negar. Muito embora o autor ache sua apresentação fundamental e tenha se divertido muito ao fazê-lo, pois a abordagem geométrica vai se desdobrando em interpretações física claras. O autor se espanta que físicos, com quem convive pelo menos, não costumam a falar de geometria com frequência e nem apresentam esse belo formalismo em cursos de base.

2 TÓPICOS ELEMENTARES EM GEOMETRIA DIFERENCIAL

Este é o tradicional capítulo dedicado a fundamentação matemática que será recorrente ao longo de todo o texto. Aqui vamos apresentar os conceitos básicos de geometria diferencial, área da matemática que nos permite tratar espaços, a primeira vista complicados, com uma estrutura localmente euclidiana suave permitindo usar ferramentas de cálculo. Nas Secs. 2.1-2.3 vamos definir os conceitos primeiros de variedade e espaços tangentes até chegarmos na Sec. 2.4 com princípios de geometria Riemanniana como conexões, derivada covariante, curvatura, torção e tensores métricos. Caso o leitor não tenha familiaridade com certas estruturas que irão aparecer nesta parte do texto, reservamos ao apêndice A alguns conceitos acerca de topologia e tensores, que são pré-requisitos para o estudo de variedades diferenciáveis.

Vale frizar que praticamente seguimos a risca o texto de Szekeres¹ neste capítulo, dado sua formalidade e acessibilidade para físicos. Pessoalmente achamos este um excelente livro de Física-Matemática moderna por tratar áreas chave da Matemática e suas projeções na Física. Apenas a discussão acerca de conexões foi retirada no livro de Amari² por questões de clareza.

2.1 Variedades diferenciáveis

Um *espaço localmente Euclidiano* ou *variedade topológica* M de dimensão n é um espaço topológico Hausdorff – com base topológica enumerável – em que cada ponto p tem uma vizinhança homeomórfica a um conjunto aberto de $\mathbb{R}^n$. Se p é ponto de M então uma *carta* em p é o par $(\mathcal{O}, \phi)$, onde $\mathcal{O}$ é um aberto de M chamado domínio da carta e $\phi : \mathcal{O} \rightarrow \phi(\mathcal{O}) \subset \mathbb{R}^n$, é um homeomorfismo entre $\mathcal{O}$ e sua imagem $\phi(\mathcal{O})$. As funções $x^i = \text{pr}_i \circ \phi : \mathcal{O} \rightarrow \mathbb{R}$ ($i = 1, \dots, n$), onde $\text{pr}_i : \mathbb{R}^n \rightarrow \mathbb{R}$ é o mapa de projeção usual, são chamadas *funções coordenadas* da carta, e os números reais $[x^i(p)]$ representam o *sistema de coordenadas* no ponto p referentes a esta carta, vide Fig. 1.

Quando quisermos enfatizar os símbolos a serem utilizados pelas funções coordenadas denotaremos a carta por $(\mathcal{O}, \phi, x^i)$, ou simplesmente $(\mathcal{O}, x^i)$. Além disso, o uso de índices sobrescrito vai se tornar claro quando apresentarmos o formalismo de tensores em variedades, a fim de utilizarmos a convenção da soma de Einstein sobre índices repetidos.

Para qualquer par de cartas $(\mathcal{O}, \phi, x^i)$ e $(\mathcal{O}', \phi', x'^j)$, tal que $\mathcal{O} \cap \mathcal{O}' \neq \emptyset$, definimos as *funções de transição*, Fig. 2, como

$$\begin{aligned}\phi' \circ \phi^{-1} &: \phi(\mathcal{O} \cap \mathcal{O}') \rightarrow \phi'(\mathcal{O} \cap \mathcal{O}') \\ \phi \circ \phi'^{-1} &: \phi'(\mathcal{O} \cap \mathcal{O}') \rightarrow \phi(\mathcal{O} \cap \mathcal{O}').\end{aligned}$$

Figura 1 – Carta $(\mathcal{O}, \phi, x^i)$ no ponto p da variedade topológica M .

Fonte: Adaptada de SZEKERES.¹

As funções de transição são também escritas como

$$x'^j = x'^j(x^1, \dots, x^n) \text{ e } x^i = x^i(x'^1, \dots, x'^n) \text{ com } i, j = 1, \dots, n, \quad (2.1)$$

que é uma forma abreviada da expressão tecnicamente correta

$$\begin{aligned} x'^j(p) &= \text{pr}_j \circ \phi' \circ \phi^{-1}(x^1(p), \dots, x^n(p)), \\ x^i(p) &= \text{pr}_i \circ \phi \circ \phi'^{-1}(x'^1(p), \dots, x'^n(p)). \end{aligned}$$

Duas cartas são ditas C^r -compatíveis para r inteiro positivo ou ∞ , se as funções de transição (2.1) são de classe C^r .

Um atlas em M é uma família de cartas $\mathcal{A} = \{(\mathcal{O}_\alpha, \phi_\alpha), \alpha \in A\}$ tal que os abertos $\mathcal{O}_\alpha$ cobrem M e qualquer par de cartas da família são C^∞ -compatíveis. Se $\mathcal{A}$ e $\mathcal{A}'$ são ambos atlas em M , então $\mathcal{A} \cup \mathcal{A}'$ também o será. Além disso, qualquer atlas $\mathcal{A}$ pode ser estendido a um atlas maximal adicionando a ele todas as cartas que são C^∞ -compatíveis com as cartas de $\mathcal{A}$. Este atlas máximo é chamado de *estrutura diferenciável ou suave* em M . Um par $(\mathcal{A}, M)$, onde M é uma variedade topológica n -dimensional e $\mathcal{A}$ é uma estrutura diferenciável em M , é chamada *variedade diferenciável ou suave*, que denotaremos apenas por M .

Figura 2 – Funções de transição entre cartas compatíveis.

Fonte: Adaptada de SZEKERES.¹

Note, a matriz Jacobiana $J = [\partial x'^k / \partial x^j]$ é não singular como evidência da estrutura diferenciável, ou seja, $\det[\partial x'^k / \partial x^j] \neq 0$, desde que

$$J^{-1}J = \begin{bmatrix} \frac{\partial x^i}{\partial x'^k} & \frac{\partial x'^k}{\partial x^j} \end{bmatrix} = \begin{bmatrix} \frac{\partial x^i}{\partial x^j} \end{bmatrix} = [\delta^i_j] = \mathbf{1},$$

onde $\mathbf{1}$ denota a matriz identidade.

2.2 Curvas e mapas diferenciáveis

2.2.1 Funções suaves e difeomorfismos

Seja M uma variedade diferencial de dimensão n . Um mapa $f : M \rightarrow \mathbb{R}$ é dito *diferenciável* no ponto $p \in M$ se para alguma carta $(\mathcal{O}, \phi, x^i)$ em p a função $\hat{f} = f \circ \phi^{-1} : \phi(\mathcal{O}) \rightarrow \mathbb{R}$ é diferenciável em $\phi(p) = \mathbf{x}(p) = (x^1(p), \dots, x^n(p))$. Esta definição é independente da escolha da carta em p , de forma que se $(\mathcal{O}', \phi')$ é uma segunda carta em p que é compatível com $(\mathcal{O}, \phi)$, então

$$\hat{f}' = f \circ \phi'^{-1} = \hat{f} \circ \phi \circ \phi'^{-1}$$

é C^∞ já que é composta de funções diferenciáveis. Denotamos por $\mathcal{F}_p(M)$ o conjunto de todas as funções de M nos reais que são diferenciáveis em $p \in M$.

Agora, dado um aberto $\mathcal{V} \subseteq M$, uma função $f : \mathcal{V} \rightarrow \mathbb{R}$ é dita *diferenciável* ou *suave* em $\mathcal{V}$ se ela é diferenciável em todo ponto $p \in \mathcal{V}$. Denotamos por $\mathcal{F}(\mathcal{V})$ o conjunto de todas as funções de M nos reais que são suaves no aberto $\mathcal{V} \subseteq M$. A soma $f + g$ e o produto fg de qualquer par de funções diferenciáveis f e g são diferenciáveis. Além do mais, $\mathcal{F}(\mathcal{V})$ é fechado por combinações lineares $f + ag$ onde $a \in \mathbb{R}$, portanto espaço vetorial comutativo com respeito a multiplicação de funções.

No que segue, sejam M e N variedades diferenciáveis de dimensão m e n respectivamente. Um mapa $\alpha : M \rightarrow N$ é diferenciável no ponto $p \in M$ se para qualquer par de cartas $(\mathcal{O}, \phi, x^i)$ e $(\mathcal{V}, \psi, y^a)$ cobrindo p e $\alpha(p)$, sua representação em termos de coordenadas

$$\hat{\alpha} = \psi \circ \alpha \circ \phi^{-1} : \phi(\mathcal{O}) \rightarrow \psi(\mathcal{V})$$

é diferenciável em $\phi(p)$. Como no caso de funções reais diferenciáveis esta definição é independente da escolha de cartas. O mapa $\hat{\alpha}$ é representado por n funções reais diferenciáveis

$$y^a = \alpha^a(x^1, \dots, x^n) \text{ com } a = 1, \dots, n,$$

onde $\alpha^a = \text{pr}_a \circ \hat{\alpha}$.

Um *difeomorfismo* é um mapa $\alpha : M \rightarrow N$ biunívoco com ambos α e $\alpha^{-1} : N \rightarrow M$ sendo diferenciáveis. Logo, duas variedades M e N são ditas *difeomorfas*, $M \cong N$, se existe um difeomorfismo $\alpha : M \rightarrow N$.

2.2.2 Curva parametrizada suave

Uma *curva parametrizada suave* em uma variedade n -dimensional M é um mapa diferenciável $\gamma : (a, b) \rightarrow M$ do aberto real $(a, b) \subseteq \mathbb{R}$ em M . Se é dito que a curva γ passa por p em $t = t_0$, então $\gamma(t_0) = p$, onde $a < t_0 < b$. Nesse sentido, fato notável é que uma mudança no parâmetro de t para $t' = f(t)$, onde $f : \mathbb{R} \rightarrow \mathbb{R}$ é uma função monotônica diferenciável e $f(a) = a' < t' < b' = f(b)$, muda a curva para $\gamma' = \gamma \circ f$, entretanto não tem efeito algum sobre a imagem em M .

Dada a carta $(\mathcal{O}, \phi, x^i)$ em p , imagem inversa do aberto $\mathcal{O}$ é o aberto $\gamma^{-1}(\mathcal{O}) \subseteq \mathbb{R}$; seja (a_1, b_1) um componente conexo deste aberto que contém o número t_0 tal que $p = \gamma(t_0)$. A representação em coordenadas da curva parametrizada γ induzida pela carta é a curva suave $\hat{\gamma} = \phi \circ \gamma : (a_1, b_1) \rightarrow \mathbb{R}^n$ descrita pelas n funções reais $x^i = \gamma^i(t)$, onde $\gamma^i = \text{pr}_i \circ \phi \circ \gamma$. Dada outra carta $(\mathcal{O}', \phi', x'^j)$ as n funções que representam a curva se tornam $x'^j = \gamma'^j(t) = x'^j(\gamma(t))$. Assumindo que as cartas são compatíveis, essas novas funções que agora representam a curva são novamente suaves, porém é perfeitamente possível que o intervalo (a', b') seja alterado.

2.3 Espaço tangente, cotangente e tensores

2.3.1 Vetores Tangentes

Seja $x^i = x^i(t)$ curva em $\mathbb{R}^n$ passando pelo ponto $\mathbf{x}_0 = \mathbf{x}(t_0)$. Em cálculo básico é comum definir a tangente à curva, ou velocidade, no ponto $\mathbf{x}_0$ como o n -vetor $\mathbf{v} = (\dot{\mathbf{x}}) = (\dot{x}^1, \dots, \dot{x}^n)$, onde $\dot{x}^i = dx^i/dt|_{t=t_0}$.

Já em uma variedade n -dimensional esta definição não é satisfatória para a direção tangente e suas componentes, visto que transformações de coordenadas gerais são permiti-

das. Assim, uma abordagem independente de coordenadas, ou *invariante*, tem motivação no conceito de *derivada direcional* de uma função diferenciável $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ao longo da curva em $\mathbf{x}_0$,

$$Df = \left. \frac{df(\mathbf{x}(t))}{dt} \right|_{t=t_0} = \left. \frac{dx^i(t)}{dt} \right|_{t=t_0} \left. \frac{\partial f(\mathbf{x})}{\partial x^i} \right|_{\mathbf{x}=\mathbf{x}_0},$$

onde D é operador linear diferencial

$$D = \left. \frac{dx^i(t)}{dt} \right|_{t=t_0} \left. \frac{\partial}{\partial x^i} \right|_{\mathbf{x}=\mathbf{x}_0}.$$

Note que o operador D quando aplicado à função f apenas depende dos valores que a função assume em uma vizinhança de $\mathbf{x}_0$ ao longo da curva em questão e é independente da escolha de coordenadas em $\mathbb{R}^n$. A expressão acima demonstra que componentes independentes de coordenadas de vetores tangentes em $\mathbb{R}^n$ podem ser extraídas dos operadores derivada direcional através dos seus coeficientes de expansão em termos derivadas parciais como coordenadas.

O *operador derivada direcional* D é um mapa real na álgebra das funções diferenciáveis em $\mathbf{x}_0$. Duas importantes propriedades valem para o mapa $D : \mathcal{F}_{\mathbf{x}_0}(\mathbb{R}^n) \rightarrow \mathbb{R}$:

- i) É linear no espaço vetorial $\mathcal{F}_{\mathbf{x}_0}(\mathbb{R}^n)$, ou seja, para qualquer par de funções f e g e números reais a e b temos $D(af + bg) = aDf + bDg$.
- ii) Aplicação de D em qualquer produto de funções fg na álgebra $\mathcal{F}_{\mathbf{x}_0}(\mathbb{R}^n)$ obedece a *regra de Leibnitz*, $D(fg) = f(\mathbf{x}_0)Dg + g(\mathbf{x}_0)Df$.

Estas propriedades motivam a definição de um vetor tangente a um ponto de uma variedade geral, além de caracterizar completamente a classe de operadores derivada direcional.

Definimos então um *vetor tangente* X_p em um ponto p da variedade diferencial M como um mapa linear $X_p : \mathcal{F}_p(M) \rightarrow \mathbb{R}$ que satisfaz:

$$X_p(af + bg) = aX_p f + bX_p g \text{ (linearidade),} \quad (2.2)$$

$$X_p(fg) = f(p)X_p g + g(p)X_p f \text{ (regra de Leibnitz).} \quad (2.3)$$

O conjunto de todos os vetores tangentes em p forma o espaço vetorial $T_p(M)$, chamado *espaço tangente em p* , já que qualquer combinação linear $aX_p + bY_p$ de vetores tangentes em p , definidas por

$$(aX_p + bY_p)f = aX_p f + bY_p f,$$

é vetor tangente em p uma vez que satisfazem (2.2) e (2.3). Dada uma carta $(\mathcal{O}, \phi, x^i)$ em p , definimos os operadores

$$(\partial_{x^i})_p \equiv \left. \frac{\partial}{\partial x^i} \right|_p : \mathcal{F}_p(M) \rightarrow \mathbb{R}$$

por

$$(\partial_{x^i})_p f \equiv \frac{\partial}{\partial x^i} \Big|_p f = \frac{\partial \hat{f}(x^1, \dots, x^n)}{\partial x^i} \Big|_{\mathbf{x}=\phi(p)}, \quad (2.4)$$

onde $\hat{f} = f \circ \phi^{-1} : \mathbb{R}^n \rightarrow \mathbb{R}$, que são claramente vetores tangentes por satisfazerem (2.2) e (2.3). Logo qualquer combinação linear da forma

$$X_p = X^i (\partial_{x^i})_p = X^i \frac{\partial}{\partial x^i} \Big|_p \equiv \sum_{i=1}^n X^i \frac{\partial}{\partial x^i} \Big|_p \text{ onde } X^i \in \mathbb{R} \quad (2.5)$$

é vetor tangente também. Os coeficientes X^j podem ser computados pela ação de X_p na função coordenada x^j de forma que

$$X_p x^j = X^i \frac{\partial x^j}{\partial x^i} \Big|_{\mathbf{x}=\phi(p)} = X^i \delta_i^j = X^j.$$

Teorema 1. Seja a carta $(\mathcal{O}, \phi, x^i)$ no ponto $p \in M$, então os operadores $(\partial_{x^i})_p$ definidos pela (2.4) formam uma base para o espaço tangente $T_p(M)$, e sua dimensão é $n = \dim M$.

A prova deste teorema está presente no texto de Szekeres.¹ Mostrando que um vetor tangente X_p tem decomposição (2.5) única com componentes $X^i = X_p x^i$ na carta $(\mathcal{O}, x^i)$.

Agora podemos relacionar a definição geral de vetor tangente com um vetor tangente a uma curva, que em $\mathbb{R}^n$ foi que nos serviu de inspiração. Seja a curva parametrizada suave $\gamma : (a, b) \rightarrow M$ passando pelo ponto $p \in M$ em $t = t_0$. Definimos o *vetor tangente a curva* $\dot{\gamma}_p$ em p pela sua ação sobre uma função diferenciável arbitrária f em p ,

$$\dot{\gamma}_p f = \frac{d}{dt} \Big|_{t=t_0} f \circ \gamma(t). \quad (2.6)$$

É direta a verificação de $\dot{\gamma}_p$ como um vetor tangente em p , observando que ele satisfaz (2.2) e (2.3). Dada a carta $(\mathcal{O}, \phi, x^i)$ em p , $\hat{\gamma} = \phi \circ \gamma = (\gamma^1(t), \dots, \gamma^n(t))$ é a representação em coordenadas da curva. Então

$$\dot{\gamma}_p f = \frac{d}{dt} \Big|_{t=t_0} f \circ \gamma(t) = \frac{\partial \hat{f}}{\partial x^i} \Big|_{\phi(p)} \frac{d\gamma^i(t)}{dt} \Big|_{t=t_0}$$

e

$$\dot{\gamma}_p = \dot{\gamma}^i(t_0) \frac{\partial}{\partial x^i} \Big|_p \text{ onde } \dot{\gamma}^i(t) = \frac{d\gamma^i(t)}{dt}.$$

No caso de $M = \mathbb{R}^n$ o operador $\dot{\gamma}_p$ é precisamente a derivada direcional da curva.

2.3.2 Espaço cotangente e tensores

O dual $T_p^*(M)$ associado ao espaço tangente em $p \in M$ é chamado *espaço cotangente* em p , composto por todos os funcionais lineares atuantes em $T_p(M)$, também chamados de

covetores ou 1-formas em p . A ação de um covetor ω_p em p sobre um vetor tangente X_p é denotada como $\omega_p(X_p)$. Em dimensão finita, vale o resultado $\dim T_p^*(M) = n = \dim T_p(M)$.

Seja f função diferenciável em p , definimos sua *diferencial* no ponto sendo o covetor $(df)_p$ que por sua vez tem ação sobre um vetor tangente X_p em p dada por

$$(df)_p(X_p) = X_p f.$$

A diferencial é um operador linear de forma que, dado X_p e Y_p vetores tangente, e escalares $a, b \in \mathbb{R}$,

$$(df)_p(aX_p + bY_p) = (aX_p + bY_p)f = aX_p f + bY_p f = a(df)_p(X_p) + b(df)_p(Y_p).$$

Assim, dada a carta $(\mathcal{O}, \phi, x^i)$ em p , as diferenciais das funções coordenadas tem a propriedade

$$(dx^i)_p(X_p) = X_p x^i = X^i$$

onde X^i são as componentes de $X_p = X^i(\partial_{x^i})_p$. Claramente atuando $(dx^i)_p$ na base do vetor tangente obtemos

$$(dx^i)_p(\partial_{x^i})_p = \frac{\partial}{\partial x^j} \Big|_p x^i = \frac{\partial x^i}{\partial x^j} \Big|_{\phi(p)} = \delta^i_j.$$

Portanto os funcionais lineares $(dx^1)_p, \dots, (dx^n)_p$ formam a base dual expandindo o espaço cotangente, assim todo covetor em p tem uma única expansão

$$\omega_p = \omega_i(dx^i)_p \text{ onde } \omega_i = \omega_p(\partial_{x^i})_p$$

são os componentes do funcional linear ω_p na carta $(\mathcal{O}, \phi, x^i)$.

A diferencial de uma função f em p tem sua expansão

$$(df)_p = f_i(dx^i)_p$$

onde

$$f_i = (df)_p(\partial_{x^i})_p = \frac{\partial}{\partial x^i} \Big|_p f = \frac{\partial \hat{f}}{\partial x^i} \Big|_{\phi(p)}.$$

Todavia, o jeito mais comum de escrever a expressão acima é

$$(df)_p = f_{,i}(dx^i)_p \text{ onde } f_{,i} = \frac{\partial \hat{f}}{\partial x^i} \circ \phi,$$

com a vírgula indicando a derivada com respeito a função coordenada x^i . As componentes $f_{,i}$ são frequentemente chamadas de *gradiente* da função em p .

Com isso, podemos agora definir um *tensor de posto* (r, s) em p como o funcional multilinear

$$A_p : \underbrace{T_p^*(M) \times \dots \times T_p^*(M)}_r \times \underbrace{T_p(M) \times \dots \times T_p(M)}_s \rightarrow \mathbb{R}.$$

Denotamos o espaço tensorial de posto (r, s) em p por $T_p^{(r,s)}(M)$, espaço vetorial de dimensão $n^{(r+s)}$.

2.3.3 Campos vetoriais e tensoriais

Um *campo vetorial* X é uma atribuição de um vetor tangente X_p em cada ponto $p \in M$, em outras palavras, X é um mapa de M no conjunto $\bigcup_{p \in M} T_p(M)$ com a propriedade de que a imagem de cada ponto, $X(p) := X_p$, pertence ao espaço tangente $T_p(M)$ em p .

O campo vetorial é dito *diferenciável* ou *suave* se para toda função diferenciável $f \in \mathcal{F}(M)$ a função Xf definida por

$$(Xf)(p) = X_p f$$

é diferenciável, ou seja, $Xf \in \mathcal{F}(M)$. O conjunto de todos os campos vetoriais em M é denotado $\mathcal{T}(M)$.

Não é de se surpreender que todo campo vetorial suave $X : \mathcal{F}(M) \rightarrow \mathcal{F}(M)$ é linear e satisfaça a regra de Leibnitz para produtos

$$X(af + bg) = aXf + bXg \text{ para toda } f, g \in \mathcal{F}(M) \text{ e para todo } a, b \in \mathbb{R},$$

$$X(fg) = fXg + gXf.$$

Reciprocamente, qualquer mapa X que satisfaça essas propriedades define um campo vetorial suave, uma vez que para cada p o mapa $X_p : \mathcal{F}(M) \rightarrow \mathcal{F}(M)$ definido por $X_p f = (Xf)(p)$ satisfaz (2.2) e (2.3) e portanto é vetor tangente em p . Também podemos definir campos vetoriais em qualquer aberto $\mathcal{O}$ de forma similar, atribuindo um vetor tangente em todo ponto de $\mathcal{O}$, tal que $Xf \in \mathcal{F}(\mathcal{O}), \forall f \in \mathcal{F}(\mathcal{O})$.

Pelo termo *base local de campos vetoriais* em p queremos dizer uma vizinhança aberta $\mathcal{O}$ de p e um conjunto de campos vetoriais $\{e_1, \dots, e_n\}$ em $\mathcal{O}$ tal que os vetores tangentes $(e_i)_q$ expandam o espaço tangente $T_q(M)$ em cada ponto $q \in \mathcal{O}$. Portanto dada a carta $(\mathcal{O}, \phi, x^i)$, definimos o campo vetorial no domínio $\mathcal{O}$

$$\partial_{x^i} \equiv \frac{\partial}{\partial x^i} : \mathcal{F}(\mathcal{O}) \rightarrow \mathcal{F}(\mathcal{O}) \text{ por } \partial_{x^i} f = \frac{\partial}{\partial x^i} f = \frac{\partial}{\partial x^i} f \circ \phi^{-1}.$$

Esses campos vetoriais conferem uma base de vetores tangentes $(\partial_{x^i})_p$ em todo ponto $p \in \mathcal{O}$, assim formam uma base local de campos vetoriais em todo ponto de $\mathcal{O}$. Quando estamos restritos ao domínio $\mathcal{O}$, todo campo vetorial X em $\mathcal{O}$ tem uma expansão única em termos dessa base

$$X|_{\mathcal{O}} = X^i \frac{\partial}{\partial x^i} = X^i \partial_{x^i},$$

onde as componentes $X^i : \mathcal{O} \rightarrow \mathbb{R}$ são funções diferenciáveis em $\mathcal{O}$. Os campos locais ∂_{x^i} formam uma base livre em $\mathcal{O}$, mas eles não são base para o espaço vetorial $\mathcal{T}(\mathcal{O})$, pois este é soma direta de espaços tangentes em todos pontos $p \in \mathcal{O}$, tendo então dimensão infinita.

De forma similar, definimos o *campo covetorial* ou *1-forma diferenciável* ω como uma atribuição de um covetor tangente ω_p em cada ponto $p \in M$, tal que a função $\omega(X)$

definida por $\omega(X)(p) = \omega_p(X_p)$ é diferenciável para todo campo vetorial suave X . O espaço das 1-formas diferenciáveis é denotado por $\mathcal{T}^*(M)$. Assim, dada qualquer função suave f , seja df a 1-forma diferencial definida pela atribuição da diferencial $(df)_p$ em cada ponto p , então

$$df(X) = Xf \text{ para todo } X \in \mathcal{T}(M).$$

Chamamos esse campo covetorial simplesmente de *diferencial* de f . Uma base livre local dada uma carta $(\mathcal{O}, \phi, x^i)$ consiste da 1-forma dx^i com a propriedade

$$dx^i(\partial_{x^j}) = \frac{\partial x^i}{\partial x^j} = \delta^i_j.$$

Toda diferencial pode ser localmente expandida como

$$df = f_{,i} dx^i \text{ onde } f_{,i} = \frac{\partial}{\partial x^i} f. \quad (2.7)$$

Campos tensoriais também são definidos de forma natural, onde um campo tensorial diferenciável A de posto (r, s) tem expansão local em uma carta qualquer

$$A = A^{i_1 \dots i_r}_{j_1 \dots j_s} \frac{\partial}{\partial x^{i_1}} \otimes \dots \otimes \frac{\partial}{\partial x^{i_r}} \otimes dx^{j_1} \otimes \dots \otimes dx^{j_s},$$

onde $\otimes$ denota o produto tensorial (vide apêndice A). As componentes são funções diferenciais sobre domínio $\mathcal{O}$ dadas por

$$A^{i_1 \dots i_r}_{j_1 \dots j_s} = A \left(dx^{i_1}, \dots, dx^{i_r}, \frac{\partial}{\partial x^{j_1}}, \dots, \frac{\partial}{\partial x^{j_s}} \right).$$

2.3.4 Transformação de coordenadas

Sejam as cartas $(\mathcal{O}, \phi, x^i)$ e $(\mathcal{O}, \phi', x'^j)$, da regra da cadeia vinda da diferenciação parcial temos

$$\frac{\partial}{\partial x'^j} = \frac{\partial x^i}{\partial x'^j} \frac{\partial}{\partial x^i}, \quad \frac{\partial}{\partial x^i} = \frac{\partial x'^j}{\partial x^i} \frac{\partial}{\partial x'^j}. \quad (2.8)$$

Substituindo as transformações (2.8) na expressão do vetor tangente com respeito a qualquer uma dessas bases temos

$$X = X^i \frac{\partial}{\partial x^i} = X'^j \frac{\partial}{\partial x'^j} \quad (2.9)$$

dando a *lei de transformação contravariante* das componentes

$$X'^j = X^i \frac{\partial x'^j}{\partial x^i}.$$

Da expressão (2.7), escrita em coordenadas x'^j e fazendo $f = x^i$, temos

$$dx^i = \frac{\partial x^i}{\partial x'^j} dx'^j.$$

Expressando uma 1-forma diferencial ω em ambos sistemas de coordenadas,

$$\omega = \omega_i dx^i = \omega'_j dx'^j,$$

obtemos a *lei de transformação covariante* das componentes

$$\omega'^j = \frac{\partial x^i}{\partial x'^j} \omega_i.$$

Portanto, a lei transformação de um tensor geral de posto (r, s) fica

$$A^{i'_1 \dots i'_r}_{j'_1 \dots j'_s} = A^{i_1 \dots i_r}_{j_1 \dots j_s} \frac{\partial x^{i'_1}}{\partial x^{i_1}} \dots \frac{\partial x^{i'_r}}{\partial x^{i_r}} \frac{\partial x^{j_1}}{\partial x^{j'_1}} \dots \frac{\partial x^{j_s}}{\partial x^{j'_s}}.$$

2.3.5 Comentário sobre fibrados

O chamado *fibrado tangente* TM a uma variedade M consiste da união dos espaços tangentes em todos os pontos da variedade

$$TM = \bigsqcup_{p \in M} T_p(M) = \bigcup_{p \in M} \{p\} \times T_p(M).$$

Podemos definir um mapa de projecção natural $\pi : TM \rightarrow M$ onde $\pi(v) = p$ para $v \in T_p(M)$ e, para cada carta $(\mathcal{O}, \phi, x^i)$ em M , definimos a carta $(\pi^{-1}(\mathcal{O}), \tilde{\phi})$ em TM onde o mapa $\tilde{\phi} : \pi^{-1}(\mathcal{O}) \rightarrow \mathbb{R}^{2n}$ é definido por

$$\begin{aligned} \tilde{\phi}(v) &= (x^1(p), \dots, x^n(p), v^1, \dots, v^n), \\ p &= \pi(v) \text{ e } v = v^i \frac{\partial}{\partial x^i} \Big|_p. \end{aligned}$$

A topologia de TM é tomada como a mais grossa tal que todos $\tilde{\phi}^{-1}(A)$ é aberto sempre que $A \subset \mathbb{R}^{2n}$ é aberto. Com esta topologia estas cartas geram um atlas maximal em TM , tornando-o uma variedade diferenciável com $\dim TM = 2n$.

Podemos dizer então que um campo vetorial em $\mathcal{O} \subseteq M$ é um mapa suave $X : \mathcal{O} \rightarrow TM$ se $\pi \circ X = id|_{\mathcal{O}}$, desde que este fato atribui um único vetor tangente de $T_p(M)$ para um ponto $p \in \mathcal{O}$. Uma ideia similar é usada para o caso de campos vetoriais suaves ao longo de uma curva $\gamma : (a, b) \rightarrow M$ definindo uma curva no fibrado $X : (a, b) \rightarrow TM$ que *levanta* γ de forma que $\pi \circ X = \gamma$. Essencialmente, isto define vetores tangentes em cada ponto da curva de forma diferenciável.

Definimos de maneira análoga o *fibrado cotangente* T^*M como a união dos espaços cotangentes em todo $p \in M$. As cartas $(\pi^{-1}(\mathcal{O}), \tilde{\phi})$ em T^*M tem seus mapas coordenados $\tilde{\phi} : \pi^{-1}(\mathcal{O}) \rightarrow \mathbb{R}^{2n}$ definidos por

$$\begin{aligned} \tilde{\phi}(\omega) &= (x^1(p), \dots, x^n(p), \omega_1, \dots, \omega_n), \\ p &= \pi(\omega) \text{ e } v = \omega_i (dx^i)_p, \end{aligned}$$

fazendo T^*M uma variedade diferenciável com $\dim T^*M = 2n$. Este processo pode ser estendido a fim de produzir um fibrado tensorial $T^{(r,s)}M$, uma variedade diferencial de dimensão $n + n^{r+s}$.

2.4 Conexões e breve introdução a geometria Riemanniana

2.4.1 Variedade pseudo-Riemanniana e Riemanniana

Um campo tensorial de posto $(0, 2)$ em uma variedade suave M é dito *não-singular* se $g_p \in T_p^{(0,2)}(M)$ é tensor não-singular em todo ponto $p \in M$. Uma *variedade pseudo-Riemanniana* (M, g) consiste de uma variedade suave M junto com um campo tensorial simétrico não-singular g de posto $(0, 2)$, chamado de *tensor métrico*. Isso é equivalente a definir uma forma não degenerada $g_p(X_p, Y_p) = \langle X_p, Y_p \rangle_p$ no espaço tangente $T_p(M)$ em todo ponto $p \in M$. Assumimos que g é campo tensorial diferenciável, então para qualquer par de vetores tangentes diferenciáveis $X, Y \in \mathcal{T}(M)$

$$g(X, Y) = g(Y, X) \in \mathcal{F}(M)$$

é uma função diferenciável. Contudo, vale notar que tal tensor métrico g , a princípio, não é naturalmente determinado pela estrutura de M como uma variedade. É possível considerar um número infinito de tensores métricos em M .

Seja a carta $(\mathcal{O}, \phi, x^i)$, então podemos escrever em forma de componentes

$$g = g_{ij} dx^i \otimes dx^j \text{ onde } g_{ij} = g_{ji} = g(\partial_{x^i}, \partial_{x^j}) = \langle \partial_{x^i}, \partial_{x^j} \rangle,$$

sendo $\mathbf{G} = [g_{ij}]$ uma matriz não-singular em cada ponto $p \in M$. Além disso, dados $X = X^i \partial_{x^i}$ e $Y = Y^i \partial_{x^i}$ em $\mathcal{T}(M)$ temos $\langle X, Y \rangle = g_{ij} X^i Y^j$.

Existe também o chamado *tensor métrico inverso* suave g^{-1} em M , um campo tensorial simétrico de posto $(2, 0)$, tal que dada uma carta como acima

$$g^{-1} = g^{ij} \frac{\partial}{\partial x^i} \otimes \frac{\partial}{\partial x^j} \text{ onde } g^{ik} g_{kj} = \delta^i_j,$$

sendo g^{ij} são as componentes da matriz inversa $\mathbf{G}^{-1}$ de $\mathbf{G}$.

Dizemos que (M, g) é *variedade Riemanniana* pedindo que o tensor métrico seja positivo definido em todo ponto

$$g_p(X_p, X_p) > 0, \text{ para todo } X_p \neq 0 \in T_p(M),$$

ou, equivalentemente, $g(X, X) \geq 0$ para todo campo vetorial $X \in \mathcal{T}(M)$. Nesse caso chamamos o campo tensorial g de posto $(0, 2)$ de *métrica Riemanniana* em M . A palavra *Riemanniana* também costuma ser empregada para se referir ao caso onde g não é positivo definido em todo ponto.

Seja $\gamma : [a, b] \rightarrow M$ uma curva parametrizada suave em uma variedade Riemanniana, o comprimento da curva entre t_0 e t é definido como

$$s = \int_{t_0}^t \sqrt{g(\dot{\gamma}(v), \dot{\gamma}(v))} dv.$$

Se a curva está contida na carta $(\mathcal{O}, x^i)$ e é escrita como $x^i = x^i(t)$ ficamos com

$$s = \int_{t_0}^t \sqrt{g_{ij} \frac{dx^i}{dv} \frac{dx^j}{dv}} dv.$$

Sendo o valor de t_0 fixo, então $ds/dt = \sqrt{g(\dot{\gamma}(t), \dot{\gamma}(t))}$ e

$$\left(\frac{ds}{dt}\right)^2 = g_{ij} \frac{dx^i}{dt} \frac{dx^j}{dt} \quad (2.10)$$

Caso a parametrização da curva seja definida com o parâmetro comprimento s , o vetor tangente é um vetor unitário ao longo da curva,

$$g(\dot{\gamma}(s), \dot{\gamma}(s)) = g_{ij} \frac{dx^i}{ds} \frac{dx^j}{ds} = 1.$$

É comum escrever a Eq.(2.10) simbolicamente na forma

$$ds^2 = g_{ij} dx^i dx^j,$$

usualmente chamada *métrica* do espaço. De fato, isto é feito para ser pensado como uma expressão simbólica que exhibe as componentes do tensor métrico e substitui a forma mais correta $g = g_{ij} dx^i \otimes dx^j$, podendo ser vista também no caso de métrica indefinida, onde temos $ds^2 < 0$.

2.4.2 Conexões afins e derivada covariante

Seja M uma variedade diferenciável n -dimensional. No caso específico em que M é um aberto de $\mathbb{R}^n$, pela definição de espaço tangente, cada espaço tangente $T_p(M)$ em cada ponto $p \in M$ pode ser visto como um $\mathbb{R}^n$, ou seja, dados pontos distintos p e q de M existe uma correspondência natural entre $T_p(M)$ e $T_q(M)$. Entretanto, no caso de M sendo uma variedade geral $T_p(M) := T_p$ e $T_q(M) := T_q$ são espaços distintos quando $p \neq q$. Desta forma, para considerarmos uma relação entre T_p e T_q , precisamos enriquecer a estrutura da variedade M . *Conexões afins* são responsáveis por esse ampliar da estrutura.

Intuitivamente, definir uma conexão afim em uma variedade M significa que para cada ponto $p \in M$ e seu “vizinho” p' , definimos um mapa linear bijetivo entre T_p e $T_{p'}$. Aqui, entendemos por ponto vizinho p' de p se, dado um sistema de coordenadas de M referente a uma carta agora denotado apenas por $[x^i]$, a diferença $dx^i \equiv x^i(p') - x^i(p)$ é suficientemente pequena de forma que podemos ignorar infinitesimais de segunda ordem $(dx^i)(dx^j)$. Queremos introduzir a noção de conexão afim utilizando infinitesimais. Poderíamos formalizar melhor a discussão usando fibrados, mas para um contexto mais físico essa abordagem é mais palpável.

Para estabelecer um mapa linear $\Pi_{p,p'} : T_p \rightarrow T_{p'}$ precisamos especificar para cada $j = 1, \dots, n$ como expressar $\Pi_{p,p'}((\partial_j)_p)$ em termos de combinações lineares de $\{(\partial_k)_{p'}\}_{k=1, \dots, n}$

$(\partial_j := \partial/\partial x^j = \partial_{x^j})$, como representado na Fig. 3. Vamos assumir que a diferença entre $\Pi_{p,p'}((\partial_j)_p)$ e $(\partial_j)_{p'}$, para p e p' vizinhos, é um infinitesimal que pode ser expresso como combinação linear de $\{dx^i\}_{i=1,\dots,n}$, ou seja,

$$\Pi_{p,p'}((\partial_j)_p) = (\partial_j)_{p'} - dx^i(\Gamma_{ij}^k)_p(\partial_k)_{p'}, \quad (2.11)$$

onde $\{(\Gamma_{ij}^k)_p; i, j, k = 1, \dots, n\}$ são n^3 números reais que dependem do ponto p .

Se, para cada par de pontos vizinhos p e p' em M é definido um mapa linear $\Pi_{p,p'}$ e se as n^3 funções $\Gamma_{ij}^k : p \rightarrow (\Gamma_{ij}^k)_p$ são todas suaves, então dizemos que introduzimos uma *conexão afim* em M . Além disso, chamamos o conjunto $\{\Gamma_{ij}^k\}$ de *coeficientes da conexão* com respeito ao sistema de coordenadas $[x^i]$. No que segue, vamos nos referir a conexões afim simplesmente por conexões.

Seja $[y^r]$ um sistema de coordenadas distinto de $[x^i]$, e seja $\tilde{\partial}_r := \partial/\partial y^r = (\partial x^i/\partial y^r)\partial_i$. Da equação (2.11) e da linearidade de $\Pi_{p,p'}$ obtemos

$$\Pi_{p,p'}((\tilde{\partial}_s)_p) = \left(\frac{\partial x^j}{\partial y^s} \right)_p \{(\partial_j)_{p'} - dx^i(\Gamma_{ij}^k)_p(\partial_k)_{p'}\}.$$

Substituindo a relação abaixo no lado direito da relação acima

$$\left(\frac{\partial x^j}{\partial y^s} \right)_{p'} = \left(\frac{\partial x^j}{\partial y^s} \right)_p + \left(\frac{\partial^2 x^j}{\partial y^r \partial y^s} \right)_p dy^r \text{ onde}$$

$$dx^i = \left(\frac{\partial x^i}{\partial y^r} \right)_p dy^r \quad \text{com} \quad dy^r \equiv y^r(p') - y^r(p),$$

ignorando termos infinitesimais de segunda ordem temos

$$\Pi_{p,p'}((\tilde{\partial}_s)_p) = (\tilde{\partial}_s)_{p'} - dy^r(\tilde{\Gamma}_{rs}^t)_p(\tilde{\partial}_t)_{p'}, \quad (2.12)$$

onde $(\tilde{\Gamma}_{rs}^t)_p$ são dados por

$$\tilde{\Gamma}_{rs}^t = \left\{ \Gamma_{ij}^k \frac{\partial x^i}{\partial y^r} \frac{\partial x^j}{\partial y^s} + \frac{\partial^2 x^k}{\partial y^r \partial y^s} \right\} \frac{\partial y^t}{\partial x^k} \quad (2.13)$$

no ponto p .

Figura 3 – Conexão afim vinda de uma abordagem usando infinitesimais.

Fonte: Adaptada de AMARI.²

Note que (2.11) e (2.12) possuem a mesma forma. Se as funções $\Gamma_{ij}^k \in C^\infty$ para todo (i, j, k) então o mesmo acontece com $\tilde{\Gamma}_{rs}^t$ para todo (r, s, t) . Desta forma, a noção de conexões afins é independente da escolha de coordenadas.

Temos então que uma conexão afim determina, dados dois pontos vizinhos p e p' , um mapeamento entre T_p e $T_{p'}$. Agora, estendendo o raciocínio conectando uma sequência de pontos vizinhos, podemos achar uma correspondência entre T_p e T_q onde p e q não são vizinhos. Esta correspondência, no entanto, depende de uma curva γ ligando p a q .

Seja $\gamma : [a, b] \rightarrow M$, onde $\gamma(a) = p$ e $\gamma(b) = q$. Um mapa de $\gamma(t)$ para um vetor tangente $X(t) \in T_{\gamma(t)}$, dito $X : t \mapsto X(t)$, é um campo vetorial ao longo da curva γ . Dado X , para todo $t \in [a, b]$ e todo infinitesimal dt queremos que os vetores tangentes correspondentes sejam linearmente relacionados como especificado pela conexão. Ou seja, se

$$X(t + dt) = \Pi_{\gamma(t), \gamma(t+dt)}(X(t)), \quad (2.14)$$

então dizemos que X é *paralelo* ao longo de γ para todo $t \in [a, b]$.

Dado o sistema de coordenadas $[x^i]$, temos que $X(t) = X^i(t)(\partial_i)_{\gamma(t)}$. Da (2.11) segue que

$$\Pi_{\gamma(t), \gamma(t+dt)}(X(t)) = \{X^k(t) - dt \dot{\gamma}^i(t) X^j(t) (\Gamma_{ij}^k)_{\gamma(t)}\} (\partial_k)_{\gamma(t+dt)},$$

onde $\gamma^i := x^i \circ \gamma$ e $\dot{\gamma}^i(t)$ é sua derivada com respeito a t . Desde que $X(t + dt) = X^k(t + dt)(\partial_k)_{\gamma(t+dt)}$, da relação (2.14), obtemos

$$\dot{X}^k(t) + \dot{\gamma}^i(t) X^j(t) (\Gamma_{ij}^k)_{\gamma(t)} = 0, \quad (2.15)$$

onde $\dot{X}^k(t) := dX^k(t)/dt = [X^k(t+dt) - X^k(t)]/dt$. A Eq.(2.15) é um sistema de equações diferenciais ordinárias em $X^1(t), \dots, X^n(t)$, assim, dado condições iniciais de contorno existe uma solução única. Segue disto que, dado $V \in T_{\gamma(a)} = T_p$ existe um único campo vetorial paralelo ao longo de γ tal que $X(a) = V$. Então fazendo $\Pi_\gamma(V)$ e denotando $X(b) \in T_{\gamma(b)} = T_q$ determinado por V vemos que Π_γ é um isomorfismo linear de T_p para T_q . Chamamos Π_γ de *transporte paralelo ao longo da curva γ* .

Seja novamente $\gamma : [a, b] \rightarrow M$ e X um campo vetorial ao longo de γ . Em geral, sabemos que $X(t)$ e $X(t+h)$ vivem em espaços tangentes distintos, portanto não é possível considerarmos coisas do tipo $dX(t)/dt = \lim_{h \rightarrow 0} [X(t+h) - X(t)]/h$. Porém, se uma conexão afim é dada em M vimos que um transporte paralelo de $X(t+h) \in T_{\gamma(t+h)}$ para $T_{\gamma(t)}$ ao longo de γ nos dá $X_t(t+h) = \Pi_{\gamma(t+h), \gamma(t)}(X(t+h))$, usando deste fato podemos considerar em $T_{\gamma(t)}$ a quantidade $\lim_{h \rightarrow 0} [X_t(t+h) - X(t)]/h$. Nós chamamos esta quantidade de *derivada covariante* de $X(t)$, denotada por $\delta X(t)/dt$. Em outras palavras, reescrevendo $X^i(t)$ como $X^i(t)(\partial_i)_{\gamma(t)}$, temos

$$\begin{aligned} \frac{\delta X(t)}{dt} &= \lim_{h \rightarrow 0} \frac{1}{h} \{X^k(t+h) - X^k(t) + [\gamma^i(t+h) - \gamma^i(t)]X^j(t+h)(\Gamma_{ij}^k)_{\gamma(t+h)}\}(\partial_k)_{\gamma(t)} \\ &= \{\dot{X}^k(t) + \dot{\gamma}^i(t)X^j(t)(\Gamma_{ij}^k)_{\gamma(t)}\}(\partial_k)_{\gamma(t)}. \end{aligned} \quad (2.16)$$

Esta expressão também corresponde a um campo vetorial ao longo de γ . Note que a condição de transporte paralelo ocorre quando $\delta X/dt = 0$.

Estender esta expressão para a derivada direcional do campo $X = X^i \partial_i \in \mathcal{T}(M)$ ao longo do vetor tangente $V = V^i(\partial_i)_p \in T_p(M)$ é direto. Considere a curva cujo vetor tangente no ponto p é V , tomando a derivada covariante de X ao longo dessa curva ficamos com

$$\nabla_V X = V^i \{(\partial_i X^k)_p + X_p^j(\Gamma_{ij}^k)_p\}(\partial_k)_p \in T_p(M). \quad (2.17)$$

De fato, fazendo $X_\gamma : t \mapsto X_{\gamma(t)}$ para qualquer curva γ , das expressões (2.16) e (2.17) temos

$$\frac{\delta X_{\gamma(t)}}{dt} = \nabla_{\dot{\gamma}(t)} X. \quad (2.18)$$

Podemos definir também para todo $X, Y \in \mathcal{T}(M)$ o campo vetorial $\nabla_X Y \in \mathcal{T}(M)$ por $(\nabla_X Y)_p = \nabla_{X_p} Y \in T_p(M)$, para todo $p \in M$. Dado que $X = X^i \partial_i$ e $Y = Y^i \partial_i$, chamamos *derivada covariante* de Y com respeito a X o campo

$$\nabla_X Y = X^i \{\partial_i Y^k + Y^j \Gamma_{ij}^k\} \partial_k. \quad (2.19)$$

Em particular, quando $X = \partial_i$ e $Y = \partial_j$ obtemos a expressão das componentes da derivada covariante

$$\nabla_{\partial_i} \partial_j = \Gamma_{ij}^k \partial_k, \quad (2.20)$$

construída como o campo vetorial que expressa a mudança do vetor ∂_j quando movido na direção ∂_i .

O operador $\nabla : \mathcal{T}(M) \times \mathcal{T}(M) \rightarrow \mathcal{T}(M)$ que mapeia $(X, Y) \mapsto \nabla_X Y$ satisfaz as seguintes propriedades para $X, Y, Z \in \mathcal{T}(M)$ e $f \in \mathcal{F}(M)$,

- i) $\nabla_{X+Y} Z = \nabla_X Z + \nabla_Y Z$.
- ii) $\nabla_X (Y + Z) = \nabla_X Y + \nabla_X Z$.
- iii) $\nabla_{fX} Y = f \nabla_X Y$.
- iv) $\nabla_X (fY) = f \nabla_X Y + (Xf)Y$.

Note que $\nabla_X Y$ é linear com respeito a X , porém não com respeito a Y , portanto ∇ não é um campo tensorial legítimo.

Nós poderíamos ter definido uma conexão afim em M como o mapa $\nabla : \mathcal{T}(M) \times \mathcal{T}(M) \rightarrow \mathcal{T}(M)$ que satisfaz as propriedades acima, onde os coeficientes $\{\Gamma_{ij}^k\}$ de ∇ com respeito a algum sistema de coordenadas $[x^i]$ seriam dados pelas n^3 funções determinadas pela Eq.(2.20). Desta forma é possível provar (2.19) e (2.13) a partir de (i)-(iv). Também é possível inverter todo o raciocínio saindo de (2.18) chegando em (2.15) voltando as definições de $\delta X/dt$ e Π_γ . Todo este processo faz desnecessário o uso de fibrados tangentes e infinitesimais, porém é bem menos intuitivo. Szekeres¹ faz exatamente isso em seu texto. A partir de agora, vamos usar a terminologia ‘conexão ∇ ’.

2.4.3 Planicidade, torção e curvatura

Vamos começar firmando um conceito. Seja $X \in \mathcal{T}$ em campo vetorial em M . Se para qualquer curva γ em M , $X_\gamma : t \mapsto X_{\gamma(t)}$ é paralelo ao longo de γ com respeito a conexão ∇ , dizemos que X é *paralelo* em M . Assim, para qualquer curva γ que conecta p e q vale $X_q = \Pi_\gamma(X_p)$. Uma condição necessária e suficiente para $X = X^i \partial_i$ ser paralelo em M é $\nabla_Y X = 0$ para todo $Y \in \mathcal{T}(M)$, ou equivalentemente

$$\partial_i X^k + X^j \Gamma_{ij}^k = 0.$$

Note que campos vetoriais paralelos em M não-nulos podem não existir em geral.

Seja $[x^i]$ em M e suponha que com respeito a ele os n vetores da base, $\{\partial_i\}_{i=1,\dots,n}$, sejam dois a dois paralelos em M . Então chamamos $[x^i]$ de *sistema de coordenadas afim* com respeito a ∇ . Esta condição é equivalente a dizer que $\nabla_{\partial_i} \partial_j = 0$ e também que os coeficientes $\{\Gamma_{ij}^k\}$ de ∇ , com respeito a $[x^i]$, são identicamente nulos. Portanto, dada uma conexão, em geral não existe um sistema de coordenadas afim correspondente. Se o sistema de coordenadas afim existe dizemos que ∇ é *plana*, ou que M é plana com respeito à ∇ .

Se $[x^i]$ é afim, então com respeito a um outro sistema de coordenadas $[y^r]$, vemos pela (2.13) que os coeficientes $\{\Gamma_{rs}^t\}$ são escritos como $\Gamma_{rs}^t = \frac{\partial^2 x^k}{\partial y^r \partial y^s} \frac{\partial y^t}{\partial x^k}$. Logo, a condição

necessária e suficiente para $[y^r]$ também ser afim é $\frac{\partial^2 x^k}{\partial y^r \partial y^s} = 0$, que é equivalente a dizer que existe uma matriz $A_{n \times n}$ e um vetor n -dimensional B tal que

$$x(p) = Ay(p) + B, \quad p \in M, \quad x(p) = [x^i(p)] \text{ e } y(p) = [y^i(p)].$$

Note que esse formato é idêntico as transformações de Poincaré da relatividade restrita. Chamamos transformações desta forma de *transformações afins* e estas são biunívocas e sua coleção forma um grupo.

Seja a conexão ∇ em M . Dados os campos vetoriais $X, Y, Z \in \mathcal{T}(M)$, definimos

$$\begin{aligned} R(X, Y)Z &\equiv \nabla_X(\nabla_Y Z) - \nabla_Y(\nabla_X Z) - \nabla_{[X, Y]}Z \text{ e} \\ T(X, Y) &\equiv \nabla_X Y - \nabla_Y X - [X, Y], \end{aligned}$$

que também são campos vetoriais de $\mathcal{T}(M)$, onde, sendo $X = X^i \partial_i$ e $Y = Y^i \partial_i$, definimos o comutador $[X, Y]$ como o campo vetorial dado por

$$[X, Y] = (X^j \partial_j Y^i - Y^j \partial_j X^i) \partial_i.$$

Os mapas $R : \mathcal{T}(M) \times \mathcal{T}(M) \times \mathcal{T}(M) \rightarrow \mathcal{T}(M)$ e $T : \mathcal{T}(M) \times \mathcal{T}(M) \rightarrow \mathcal{T}(M)$ como definidos acima são ambos $\mathcal{F}(M)$ -multilineares e campos tensoriais de posto $(1, 3)$ e $(1, 2)$, respectivamente. Chamamos R de *tensor de curvatura* e T de *tensor de torção* de ∇ . Uma motivação mais rigorosa para a definição destes tensores pode ser vistas também em Szekeres.¹

A expressão em componentes de R e T com respeito a um sistema de coordenadas $[x^i]$ são dadas por $R(\partial_i, \partial_j) \partial_k = R_{ijk}^l \partial_l$ e $T(\partial_i, \partial_j) = T_{ij}^k \partial_k$, e podem ser descritas da seguinte forma

$$\begin{aligned} R_{ijk}^l &= \partial_i \Gamma_{jk}^l - \partial_j \Gamma_{ik}^l + \Gamma_{jh}^l \Gamma_{jk}^h - \Gamma_{jh}^l \Gamma_{ik}^h \text{ e} \\ T_{ij}^k &= \Gamma_{ij}^k - \Gamma_{ji}^k. \end{aligned}$$

Claramente, se ∇ é plana então $R = 0$ e $T = 0$. No caso da volta, se $R = 0$ e $T = 0$, ∇ é localmente plana no sentido que, para cada $p \in M$ existe uma vizinhança aberta $\mathcal{O}$ de p tal que ∇ é plana em $\mathcal{O}$.³⁰

Das expressões das componentes dos tensores de curvatura e torção já podemos ver algumas de suas simetrias, $R_{ijk}^l = -R_{jki}^l$ e $T_{ij}^k = -T_{ji}^k$, portanto no caso particular de M unidimensional $R = T = 0$ vale, assim M é plana. Em geral, quando $T = 0$ vale ($\Gamma_{ij}^k = \Gamma_{ji}^k$) chamamos ∇ de *conexão livre de torção* ou *conexão simétrica*.

2.4.4 Conexão Riemanniana

Antes de iniciar a discussão pretendida aqui vamos apresentar três resultados úteis e necessários para o conteúdo. O primeiro é que as componentes da derivada covariante

de um campo tensorial $T \in \mathcal{T}^{(r,s)}(M)$ com respeito ao campo vetorial $X = X^p \partial_p \in \mathcal{T}(M)$ são dadas por

$$\begin{aligned} \nabla_X T^{ij..}_{kl..} = & X^p \{ \partial_p T^{ij..}_{kl..} + \Gamma_{ap}^i T^{aj..}_{kl..} + \Gamma_{ap}^j T^{ia..}_{kl..} + \dots \\ & - \Gamma_{kp}^a T^{ij..}_{al..} - \Gamma_{lp}^a T^{ij..}_{ka..} - \dots \}. \end{aligned}$$

O segundo é que para todo campo tensorial $T \in \mathcal{T}^{(r,s)}(M)$, sua derivada covariante ∇T é o campo tensorial de posto $(r, s+1)$ definido por

$$\nabla T(\omega^1, \dots, \omega^r, Y_1, \dots, Y_s, X) = \nabla_X T(\omega^1, \dots, \omega^r, Y_1, \dots, Y_s).$$

Por fim, o terceiro é que

$$\begin{aligned} X(T(\omega^1, \dots, \omega^r, Y_1, \dots, Y_s)) = & \nabla_X T(\omega^1, \dots, \omega^r, Y_1, \dots, Y_s) \\ & + T(\nabla_X \omega^1, \dots, \omega^r, Y_1, \dots, Y_s) + \dots + T(\omega^1, \dots, \nabla_X \omega^r, Y_1, \dots, Y_s) \\ & + T(\omega^1, \dots, \omega^r, \nabla_X Y_1, \dots, Y_s) + \dots + T(\omega^1, \dots, \omega^r, Y_1, \dots, \nabla_X Y_s). \end{aligned}$$

Esses resultados tem suas derivações feitas também em Szekeres.¹

Uma variedade pseudo-Riemanniana $(M, g = \langle \cdot, \cdot \rangle)$ possui uma *conexão natural* em si, ∇ , que satisfaz os seguintes requisitos:

- i) ∇ é livre de torção.
- ii) $\nabla g = 0$.

Esta conexão é chamada *conexão Riemanniana* definida pelo tensor métrico g . Dado um tensor métrico g essa conexão existe e é única.

Olhando para a consequência da condição ii). Seja γ a curva com vetor tangente $Z(t) = \dot{\gamma}(t)$ e X, Y campos vetoriais paralelamente transportados ao longo de γ , ou seja, $\nabla_Z X = \nabla_Z Y = 0$, segue que $g(X, Y)$ é constante ao longo da curva

$$\begin{aligned} \frac{d}{dt} g(X, Y) &= \frac{\delta}{dt} g(X, Y) = \nabla_Z (g(X, Y)) \\ &= (\nabla_Z g)(X, Y) + g(\nabla_Z X, Y) + g(X, \nabla_Z Y) \\ &= \nabla g(Z, X, Y) = 0. \end{aligned}$$

Em particular, todo campo vetorial Y transportado paralelamente ao longo da curva γ tem magnitude constante $g(Y, Y)$ ao longo da mesma, uma condição que é, de fato, necessária e suficiente para valer ii).

Uma outra forma de escrever a condição ii), onde agora $X, Y, Z \in \mathcal{T}(M)$ são quaisquer, é

$$Z \langle X, Y \rangle = \langle \nabla_Z X, Y \rangle + \langle X, \nabla_Z Y \rangle. \quad (2.21)$$

Agora vamos provar a unicidade e a expressão da conexão Riemanniana em termos do tensor métrico g . Sejam as coordenadas $[x^i]$, a condição i) nos dá que $\Gamma_{jk}^i = \Gamma_{kj}^i$, a condição ii) pode ser reescrita como

$$\begin{aligned}\nabla_k g_{ij} &= \partial_k g_{ij} - \Gamma_{ik}^m g_{mj} - \Gamma_{jk}^m g_{im} = 0 \\ g_{ij;k} &= g_{ij,k} - \Gamma_{ik}^m g_{mj} - \Gamma_{jk}^m g_{im} = 0,\end{aligned}\tag{2.22}$$

onde $g_{ij;k} := \nabla_k g_{ij} = \nabla_{\partial_k} g_{ij}$ e $g_{ij,k} := \partial_k g_{ij}$.

Intercalando pares de índices i, k e j, k obtemos

$$g_{kj;i} = g_{kj,i} - \Gamma_{ki}^m g_{mj} - \Gamma_{ji}^m g_{km} = 0,\tag{2.23}$$

$$g_{ik;j} = g_{ik,j} - \Gamma_{ij}^m g_{mk} - \Gamma_{kj}^m g_{im} = 0.\tag{2.24}$$

Usando a simetria de g_{ij} e Γ_{ij}^m , combinamos as Eqs.(2.22)+(2.23)-(2.24) e chegamos em em

$$g_{ij,k} + g_{kj,i} - g_{ik,j} = 2\Gamma_{ik}^m.$$

Multiplicando a expressão acima por g^{jl} , depois de uma mudança de índices ($l \rightarrow i$), ($i \rightarrow j$), ($j \rightarrow m$), temos

$$\Gamma_{jk}^i = \frac{1}{2}g^{im}(g_{mj,k} + g_{mk,j} - g_{jk,m}).\tag{2.25}$$

Estes coeficientes são chamados de *símbolos de Christoffel*, eles são a forma explícita das componentes da conexão Riemanniana escritos em termos do tensor métrico em qualquer sistema de coordenadas.

3 GEOMETRIA DE MODELOS ESTATÍSTICOS CLÁSSICOS E QUÂNTICOS

Neste capítulo, vamos estudar aspectos geométricos que emanam de espaços de estados clássicos e quânticos. Vale destacar que para isso será necessário ter em mente o formalismo apresentado no capítulo anterior. Esta discussão é feita de maneira muito completa por Amari,² o qual seguiremos a risca, fazendo algumas adaptações, executando grande parte das demonstrações e interpretando resultados centrais que aparecerão direta ou indiretamente em discussões futuras.

Na Sec. 3.1 definiremos o chamado modelo estatístico $\mathcal{S}$, espaço de distribuições de probabilidade que tem estrutura de variedade diferenciável, base para nossa abordagem. As Sec. 3.2-3.5 são dedicadas ao espaço de estados clássico, onde temos a métrica de Fisher g como única invariante sob estatística suficiente, definimos as α -conexões que apresentam uma estrutura dual com respeito a métrica de Fisher, discutimos sobre a origem afim destas conexões e construímos a estrutura dualística g , $\pm\alpha$ -conexões através das α -divergências globalmente definidas em $\mathcal{S}$, de forma que essas divergências correspondem a uma série de entropias relativas já conhecidas na literatura.

A Sec. 3.6 é reservada a aplicarmos os procedimentos desenvolvidas no caso clássico ao caso quântico, definindo através de α -divergências quânticas um análogo formal a estrutura Fisher e $\pm\alpha$ -conexões. Observamos que o espaço de estados quânticos se quebra em uma direção comutativa isomorfa ao caso clássico e uma parte não-comutativa que tem uma família α de métricas associada a maneira que mergulhamos nossa variedade.

No apêndice B temos uma apresentação, mais familiar para físicos, que discute conceitos acerca de metrologia clássica e quântica, tendo uma motivação da definição da informação de Fisher e a derivação da cota de Cramér-Rao, esta assunto de capítulo futuro. Pode ser uma leitura muito instrutiva, caso seja o primeiro contato do leitor com estes tópicos, recomendamos sua leitura antes de iniciar este capítulo se for este o caso. Já o apêndice C traz uma discussão geométrica do cota de Cramér-Rao (CRB)* que decorre naturalmente deste capítulo, recomendamos muito sua leitura após terminar este capítulo.

3.1 Modelos estatísticos

Seja $\mathcal{X}$ um conjunto com seus elementos sendo m -uplas reais, por simplicidade. Uma *distribuição de probabilidade* em $\mathcal{X}$, supondo aqui que ele seja contável, é uma função $p : \mathcal{X} \rightarrow \mathbb{R}$ que satisfaz

$$p(x) \geq 0 \quad (\forall x \in \mathcal{X}) \text{ e } \sum_{x \in \mathcal{X}} p(x) = 1.$$

*O leitor irá notar que vamos utilizar ao longo do texto a sigla CRB, do inglês *Cramér-Rao bound*, para nos referirmos a cota de Cramér-Rao.

Se $\mathcal{X} = \mathbb{R}^n$, então a função p deve satisfazer

$$p(x) \geq 0 \quad (\forall x \in \mathcal{X}) \text{ e } \int p(x)dx = 1, \quad (3.1)$$

em outras palavras, p é função densidade de probabilidade em $\mathcal{X}$.

Vale mencionar que um tratamento estatístico-matemático mais rigoroso considera a função densidade de probabilidade como $p = dP/d\nu : \mathcal{X} \rightarrow \mathbb{R}$, onde ν é uma σ -medida finita em um espaço mensurável $(\mathcal{X}, \mathcal{B})$, com $\mathcal{B}$ sendo uma σ -álgebra (classe completamente aditiva) consistindo de $\mathcal{X}$ e seus subconjuntos, P é função probabilidade em $(\mathcal{X}, \mathcal{B})$ que é contínua com respeito a ν . O leitor interessado pode ver as referências em teoria de probabilidade³¹⁻³⁴ se desejar mais familiaridade com essa teoria.

Utilizaremos a expressão integral (3.1) ao longo da discussão. Euristicamente, os resultados para o caso discreto podem ser obtidos substituindo a integral $\int \cdots dx$ pela soma $\sum_{x \in \mathcal{X}} \cdots$.

Considere uma família $\mathcal{S}$ de distribuições de probabilidade em $\mathcal{X}$. Suponha que cada elemento de $\mathcal{S}$ é parametrizado por n variáveis reais $(\theta^1, \dots, \theta^n)$, de forma que

$$\mathcal{S} = \{p_\theta = p(x, \theta) \mid \theta = (\theta^1, \dots, \theta^n) \in \Theta\},$$

onde $\Theta \subseteq \mathbb{R}^n$ e o mapa $\theta \mapsto p_\theta$ é injetivo. Chamamos $\mathcal{S}$ de *modelo estatístico*, *modelo paramétrico* ou simplesmente *modelo n -dimensional* em $\mathcal{X}$. De forma abreviada podemos denotar $\mathcal{S} = \{p(x, \theta)\} = \{p_\theta\}$.

Vamos assumir que Θ é um aberto de $\mathbb{R}^n$ de forma que para cada $x \in \mathcal{X}$ a função $\theta \mapsto p_\theta$ é C^∞ . Isto permite que expressões como $\partial_i p_\theta$ e $\partial_i \partial_j p_\theta = \partial_j \partial_i p_\theta$ sejam bem definidas, onde $\partial_i \equiv \partial/\partial \theta^i$. Além disso

$$\int \partial_i p_\theta dx = \partial_i \int p_\theta dx = \partial_i 1 = 0. \quad (3.2)$$

Para a distribuição de probabilidade p em $\mathcal{X}$, dizemos que o suporte de p é dado por $\text{supp}(p) \equiv \{x \in \mathcal{X} \mid p(x) > 0\}$. Vamos considerar apenas o caso em que $\text{supp}(p_\theta)$ é constante com respeito a θ , o outro caso traz dificuldade desnecessária à discussão. Seja $\mathcal{X}$ redefinido como $\text{supp}(p_\theta)$, equivalente a assumir que $p_\theta > 0$ para todo $\theta \in \Theta$ e $x \in \mathcal{X}$, isto significa que $\mathcal{S}$ é subconjunto de

$$\mathcal{P}(\mathcal{X}) \equiv \left\{ p : \mathcal{X} \rightarrow \mathbb{R} \mid p(x) > 0 \quad (\forall x \in \mathcal{X}), \int p(x)dx = 1 \right\}.$$

Dado um modelo estatístico $\mathcal{S} = \{p_\theta\}$, o mapa $\phi : \mathcal{S} \rightarrow \mathbb{R}^n$ definido por $\phi(p_\theta) = \theta$ nos fornece um sistema de coordenadas $[\theta^i]$ de $\mathcal{S}$. Suponha então que temos um difeomorfismo $\psi : \Theta \rightarrow \psi(\Theta)$. Então, se $\rho = \psi(\theta)$, obtemos $\mathcal{S} = \{p_{\psi^{-1}(\rho)} \mid \rho \in \psi(\Theta)\}$, onde esta expressão representa a mesma família de distribuições de probabilidade $\mathcal{S} = \{p_\theta\}$.

Se considerarmos parametrizações que são difeomorfas entre si como equivalentes, podemos considerar $\mathcal{S}$ como uma variedade diferenciável suave chamada *variedade estatística*. Claramente, neste caso a parametrização de $\mathcal{S}$ é o próprio sistema de coordenadas.

3.2 Geometria de modelos clássicos, métrica de Fisher e α -conexões

3.2.1 Matriz informação de Fisher

Seja o modelo estatístico $\mathcal{S} = \{p_\theta \mid \theta \in \Theta \subset \mathbb{R}^n\} \subset \mathcal{P}(\mathcal{X})$. Dado um ponto $\theta \in \Theta$ a *matriz informação de Fisher* de $\mathcal{S}$ em θ é a matriz $n \times n$ simétrica positiva $\mathbf{G}(\theta) = [g_{ij}(\theta)]$ que tem seus componentes definidos por

$$g_{ij}(\theta) \equiv \int \partial_i l(x, \theta) \partial_j l(x, \theta) p(x, \theta) dx = E_\theta[\partial_i l_\theta \partial_j l_\theta], \quad (3.3)$$

onde $l(x, \theta) \equiv \log[p(x, \theta)] = l_\theta$ e $E_\theta[f] \equiv \int f(x) p(x, \theta) dx$ é o valor esperado com respeito a distribuição p_θ . A função $\log$ aqui, denota logaritmo na base natural.

Para o caso de $n = 1$, chamamos a grandeza de *informação de Fisher*. Note que da expressão (3.2) podemos concluir que

$$E_\theta[\partial_i l_\theta] = 0 \xrightarrow{\partial_j} -E_\theta[\partial_i \partial_j l_\theta] = g_{ij}(\theta). \quad (3.4)$$

Uma outra representação muito importante para as componentes é

$$g_{ij}(\theta) = 4 \int \partial_i \sqrt{p_\theta} \partial_j \sqrt{p_\theta} dx, \quad (3.5)$$

que irá aparecer naturalmente no Cap. 5.

Podemos então definir um tensor métrico em $\mathcal{S}$, dado o sistema de coordenadas $[\theta^i]$, de forma que $g_{ij} = \langle \partial_i, \partial_j \rangle$. Assim determinamos unicamente uma métrica Riemanniana $g = \langle \cdot, \cdot \rangle$. Esta é a chamada *métrica de Fisher* ou *métrica da informação*, que é uma métrica privilegiada sobre a ação de estatísticas suficientes pelo Teo. 6 de Chentsov. Além disso, é dito na introdução do texto de Jenčova³⁵ que, a métrica de Fisher é a única métrica contrativa sob ação de mapa Markoviano, o que a tornaria uma métrica única para o espaço de estados clássico.

Desde que a métrica g_{ij} dada por (3.3) se transforma como um tensor legítimo, assim vemos que a métrica de Fisher é covariante sobre escolha de sistema de coordenadas. De fato, podemos escrever $\langle X, Y \rangle_\theta = E_\theta[(X l_\theta)(Y l_\theta)]$ para todo vetor tangente $X, Y \in T_\theta(\mathcal{S})$. Vamos omitir o sub-índice θ em $X_\theta \in T_\theta(\mathcal{S})$ quando preciso para evitar carregar a notação e redundâncias como $X_\theta p_\theta$, por exemplo.

3.2.2 Estatística suficiente

Dada $F : \mathcal{X} \rightarrow \mathcal{Y}$ que transforma valores de x em $F(x) = y$, definimos $q(y, \theta) \equiv p(x, \theta)$ tal que $y = F(x)$.

Apresentando um pouco de notação, temos

$$\begin{aligned} r(x, \theta) &= \frac{p(x, \theta)}{q(F(x), \theta)} \\ p(x \mid y, \theta) &= r(x, \theta) \delta_{F(x)}(y) \\ \Pr(\mathcal{O} \mid y, \theta) &= \int_{\mathcal{O}} p(x \mid y, \theta) dx, \quad \mathcal{O} \subset \mathcal{X}, \end{aligned}$$

onde $\delta_{F(x)}$ é a função delta de Dirac em $\mathcal{Y}$ no ponto $F(x) = y$. Para qualquer $\mathcal{V} \subset \mathcal{Y}$

$$\int_{\mathcal{O} \cap F^{-1}(\mathcal{V})} p(x, \theta) dx = \int_{\mathcal{V}} \Pr(\mathcal{O} \mid y, \theta) q(y, \theta) dy,$$

ou seja, $\Pr(\mathcal{O} \mid y, \theta)$ é a probabilidade condicional de um evento $x \in \mathcal{O}$ dado y . Caso $\Pr(\mathcal{O} \mid y, \theta)$ não dependa de θ para todo $\mathcal{O}$ e y , ou equivalentemente, se $r(x, \theta)$ não depende de θ para todo x , então dizemos que F é *estatística suficiente* para o modelo $\mathcal{S}$. Mapeamentos injetores são exemplos triviais de estatísticas suficientes.

Se F é estatística suficiente podemos escrever que

$$p(x, \theta) = q(F(x), \theta) r(x),$$

ou seja, a parte de $p(x, \theta)$ que depende de θ está toda contida na distribuição $q(y, \theta)$ de $y = F(x)$. Portanto, para estimar o parâmetro desconhecido θ basta conhecermos os valores de y . Afinal, dado um valor de y podemos simular x usando $p(x \mid y)$ que não depende de θ .

Teorema 2. A matriz informação de Fisher $G_F(\theta) = [g_{ij}^F(\theta)]$ do modelo induzido $\mathcal{S}_F \equiv \{q(y, \theta)\}$ satisfaz $G_F(\theta) \leq G(\theta)$ (monotonicidade), onde $G(\theta) = [g_{ij}(\theta)]$ é a matriz informação de Fisher do modelo original $\mathcal{S}$, de forma que $\Delta G(\theta) \equiv G(\theta) - G_F(\theta)$ é positiva semi-definida. A condição necessária e suficiente para termos $G_F(\theta) = G(\theta)$ é que F seja estatística suficiente para $\mathcal{S}$. Em geral, a perda de informação $\Delta G(\theta) = [\Delta g_{ij}(\theta)]$ causada por codificar x em $y = F(x)$ é dada por

$$\Delta g_{ij}(\theta) = E_{\theta}[\partial_i \log r(x, \theta) \partial_j \log r(x, \theta)].$$

Geometricamente falando, uma interpretação que emana do Teorema acima é que F ser estatística suficiente para $\mathcal{S}$ é a mesma coisa que ela ser isometria com respeito a métrica de Fisher.

A demonstração desse teorema não acrescenta muito a nossa discussão, caso o leitor se interesse, veja o texto de Amari.²

3.2.3 As α -conexões

Considere as funções $\Gamma_{kij}^{(\alpha)}$ que mapeiam θ para o seguinte valor

$$\left(\Gamma_{kij}^{(\alpha)}\right)_{\theta} \equiv E_{\theta} \left[\left(\partial_i \partial_j l_{\theta} + \frac{1-\alpha}{2} \partial_i l_{\theta} \partial_j l_{\theta} \right) (\partial_k l_{\theta}) \right], \quad (3.6)$$

onde $\alpha \in \mathbb{R}$. As n^3 funções $\Gamma_{kij}^{(\alpha)}$, sobre uma mudança de coordenadas, se transformam como a Eq.(2.13). Assim temos uma conexão simétrica afim $\nabla^{(\alpha)}$ em $\mathcal{S}$ definida por

$$\langle \nabla_{\partial_i}^{(\alpha)} \partial_j, \partial_k \rangle = \Gamma_{kij}^{(\alpha)},$$

onde, claramente, $g = \langle \cdot, \cdot \rangle$ é a métrica de Fisher. Nós chamamos $\nabla^{(\alpha)}$ de α -conexão.

Em particular vamos olhar para o caso em que $\alpha = 0$. Temos que

$$\partial_j l_\theta = \frac{1}{p_\theta} \partial_j p_\theta \quad \text{e} \quad \partial_i \partial_j l_\theta = -\frac{1}{p_\theta^2} \partial_i p_\theta \partial_j p_\theta + \frac{1}{p_\theta} \partial_i \partial_j p_\theta.$$

Substituindo na expressão (3.6), chegamos que para $\alpha = 0$

$$\Gamma_{kij}^{(0)} = \int dx \left(\frac{1}{p_\theta} \partial_k p_\theta \partial_i \partial_j p_\theta - \frac{1}{2p_\theta^2} \partial_i p_\theta \partial_j p_\theta \partial_k p_\theta \right), \quad (3.7)$$

expressão que vai reaparecer também no Cap. 5.

Tomando a derivada parcial de g_{ij} na expressão (3.3) com respeito a θ^k obtemos

$$\begin{aligned} \partial_k g_{ij} &= E_\theta[\partial_k \partial_i l_\theta \partial_j l_\theta] + E_\theta[\partial_i l_\theta \partial_k \partial_j l_\theta] + E_\theta[\partial_i l_\theta \partial_j l_\theta \partial_k l_\theta] \\ &= \Gamma_{jki}^{(0)} + \Gamma_{ikj}^{(0)}, \end{aligned} \quad (3.8)$$

que é exatamente a expressão (2.22) da discussão sobre conexões Riemannianas.

Teorema 3. A 0-conexão é uma conexão Riemanniana com respeito a métrica de Fisher.

Em geral, quando $\alpha \neq 0$, $\nabla^{(\alpha)}$ não é Riemanniana.

3.3 Discussão sobre a origem afim das α -conexões

Anteriormente trabalhamos com $\mathcal{S} = \{p_\theta\} \subset \mathcal{P}(\mathcal{X})$, onde cada p_θ representa uma distribuição de probabilidades. Agora, vamos relaxar a normalização, contudo mantendo-a finita. Então, $\mathcal{S} = \{p_\theta\}$ vai ser subconjunto de

$$\tilde{\mathcal{P}}(\mathcal{X}) \equiv \left\{ p : \mathcal{X} \rightarrow \mathbb{R} \mid p(x) > 0 \quad (\forall x \in \mathcal{X}), \int p(x) dx < \infty \right\}.$$

No que segue, mantemos as afirmações feitas nas seções anteriores, Sec. 3.1 e Sec. 3.2, de forma que $\mathcal{S}$ tem estrutura de variedade juntamente com a métrica de Fisher e as α -conexões definidas usando as equações (3.3) e (3.6), respectivamente. Entretanto, veremos que as expressões (3.4) não valerão mais em geral dado o mergulho que vamos definir a seguir.

Para cada $\alpha \in \mathbb{R}$ definimos

$$\begin{aligned} L^{(\alpha)}(u) &\equiv \begin{cases} \frac{2}{1-\alpha} u^{(\frac{1-\alpha}{2})}, & (\alpha \neq 1) \\ \log(u), & (\alpha = 1) \end{cases} \\ \ell_{p_\theta}^{(\alpha)} &\equiv L^{(\alpha)}(p_\theta) \end{aligned} \quad (3.9)$$

Identificando $\mathcal{S}$ com o conjunto $\{\ell_{p_\theta}^{(\alpha)} \mid p_\theta \in \mathcal{S}\}$. Em particular $\ell^{(1)} = l_\theta$ e $\ell^{(-1)} = p_\theta$, denotando $\ell_{p_\theta}^{(\alpha)}$ apenas por $\ell^{(\alpha)}$

Desta forma, um vetor tangente $X \in T_\theta(\mathcal{S})$ deve valer

$$X^{(\alpha)}(x) \equiv X\ell^{(\alpha)},$$

sendo uma função de x a θ fixo, esta é a α -representação de X_θ .

Desde que

$$\partial_i \ell^{(\alpha)} = p_\theta^{(1-\alpha)/2} \partial_i l_\theta \quad \text{e} \quad \partial_i \partial_j \ell^{(\alpha)} = p_\theta^{(1-\alpha)/2} \left(\partial_i \partial_j l_\theta + \frac{1-\alpha}{2} \partial_i l_\theta \partial_j l_\theta \right),$$

podemos reescrever as equações (3.3) e (3.6) como

$$g_{ij}(\theta) = \int \partial_i \ell^{(\alpha)} \partial_j \ell^{(-\alpha)} dx, \quad (3.10)$$

$$\Gamma_{kij}^{(\alpha)}(\theta) = \int \partial_i \partial_j \ell^{(\alpha)} \partial_k \ell^{(-\alpha)} dx. \quad (3.11)$$

Essas equações nos mostram que a métrica de Fisher e as α -conexões em $\mathcal{S}$ são induzidas da estrutura afim do espaço de funções $\mathbb{R}^{\mathcal{X}} = \{A \mid A : \mathcal{X} \rightarrow \mathbb{R}\} \supset \tilde{\mathcal{P}}(\mathcal{X})$ através do mergulho $p \mapsto \ell^{(\alpha)}$. Em outras palavras, a origem afim das α -conexões é herdada de um espaço mais simples no qual a variedade $\mathcal{S}$ foi mergulhada, por isso a métrica de Fisher tem uma aparência um tanto Euclidiana.

O mergulho (3.9) permite colocarmos $\mathcal{S}$ dentro de um “ $\mathbb{R}^n$ ” de forma que as α -conexões são induzidas por ele via estrutura de variedade deste espaço de fundo. Esse “ $\mathbb{R}^n$ ” tem uma métrica e uma conexão Riemanniana natural em si. A componente tangente desta conexão em $\mathcal{S}$ mergulhado são as α -conexões, para algum valor α , que não são necessariamente serão Riemannianas com respeito à Métrica de Fisher em $\mathcal{S}$.

Entretanto temos que tomar cuidado com a dimensão infinita de $\mathbb{R}^{\mathcal{X}}$, afinal $\mathcal{X}$ infinito faz com que $\tilde{\mathcal{P}}(\mathcal{X})$ não seja variedade, por isso escrevemos acima $\mathbb{R}^n$ entre aspas. No apêndice C comentamos brevemente este fato. Note que se $\mathcal{S}$ está no subespaço onde as distribuições de probabilidade são normalizadas, podemos reescrever as expressões (3.4) como

$$\int p_\theta^{(\frac{1+\alpha}{2})} \partial_i \ell^{(\alpha)} dx = 0,$$

$$\frac{1+\alpha}{2} g_{ij}(\theta) = - \int p_\theta^{(\frac{1+\alpha}{2})} \partial_i \partial_j \ell^{(\alpha)} dx.$$

3.4 Conexões Duais

Seja o par $(\mathcal{S}, g)$, onde $g = \langle \cdot, \cdot \rangle$ é métrica Riemanniana em $\mathcal{S}$, e duas conexões afins ∇ e ∇^* . Se para quaisquer $X, Y, Z \in \mathcal{T}(\mathcal{S})$ vale

$$Z \langle X, Y \rangle = \langle \nabla_Z X, Y \rangle + \langle X, \nabla_Z^* Y \rangle, \quad (3.12)$$

então ∇ e ∇^* são ditas *duais* entre si com respeito a g , ou ainda, uma é *conexão dual* ou *conexão conjugada* da outra. A tripla (g, ∇, ∇^*) tem o nome de *estrutura dualística* em $\mathcal{S}$. Com respeito a um sistema de coordenadas $[\theta^i]$ podemos reescrever a expressão (3.12) como

$$g_{ij,k} = \Gamma_{jik} + \Gamma_{ikj}^*, \quad (3.13)$$

onde g_{ij} , Γ_{kij} e Γ_{kij}^* são as componentes de g , ∇ e ∇^* . Vale salientar que $(\nabla^*)^* = \nabla$ é válida.

Dada uma métrica g e uma conexão ∇ em $\mathcal{S}$ existe uma única conexão dual ∇^* de ∇ com respeito a g . Além disso, temos que $(\nabla + \nabla^*)/2$ é conexão Riemanniana, onde a recíproca é se a conexão ∇' tendo a mesma torção que ∇^* e se $(\nabla + \nabla')/2$ é Riemanniana então $\nabla' = \nabla^*$.

Teorema 4. Para qualquer modelo estatístico $\mathcal{S}$, a α -conexão e a $(-\alpha)$ -conexão são duais com respeito a métrica de Fisher.

Prova. A demonstração segue de forma imediata da equação (3.8).

$$\begin{aligned} \partial_k g_{ij} &= E_\theta[\partial_k \partial_i l_\theta \partial_j l_\theta] + E_\theta[\partial_i l_\theta \partial_k \partial_j l_\theta] + \left(1 + \frac{\alpha - \alpha}{2}\right) E_\theta[\partial_i l_\theta \partial_j l_\theta \partial_k l_\theta], \quad \alpha \in \mathbb{R} \\ &= \Gamma_{jki}^{(\alpha)} + \Gamma_{ikj}^{(-\alpha)}. \end{aligned}$$

Uma outra forma de demonstrar esse fato é derivando a expressão (3.10) e utilizando a definição de (3.11). $\square$

É nítida a semelhança entre (3.12), (3.13) com as expressões (2.21) e (2.22), respectivamente. A condição para uma conexão ∇ ser Riemanniana é equivalente a requerer que $\nabla = \nabla^*$ na variedade, de forma que podemos considerar a dualidade entre conexões sendo uma generalização da noção de conexão Riemanniana. No texto de Amari e Nagaoka, existe uma breve discussão sobre este fato.²

3.5 α -divergências em modelos estatísticos

3.5.1 Divergências

Nesta seção queremos mostrar que existe uma relação entre uma função divergência D , definida em $\mathcal{S}$, com a estrutura dualística $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$.

Seja $\mathcal{S}$ variedade possuindo uma função suave $D(\cdot \parallel \cdot) : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}$ satisfazendo para quaisquer dois pontos $p, q \in \mathcal{S}$

$$D(p \parallel q) \geq 0 \quad \text{e} \quad D(p \parallel q) = 0 \text{ implica } p = q. \quad (3.14)$$

A função $D(\cdot \parallel \cdot)$ pode parecer algum tipo de medida de distância entre dois pontos. Entretanto não satisfaz em geral os outros axiomas de métrica (simetria e desigualdade triangular).

Dado um sistema de coordenadas $[\theta^i]$ de $\mathcal{S}$, vamos representar o par de pontos $(p, q) \in \mathcal{S} \times \mathcal{S}$ pelo par de coordenadas (θ^i, θ^i) e denotar derivadas parciais de $D(p \parallel q)$ com respeito a p e q por

$$\begin{aligned} D((\partial_i)_p \parallel q) &\equiv \partial_i D(p \parallel q) \\ D((\partial_i)_p \parallel (\partial_j)_q) &\equiv \partial_i \partial'_j D(p \parallel q) \\ D((\partial_i \partial_j)_p \parallel (\partial_k)_q) &\equiv \partial_i \partial_j \partial'_k D(p \parallel q), \text{ etc.}, \end{aligned}$$

onde no lado direito de cada expressão acima ∂_i é aplicada a primeira variável de D em p e ∂'_i na segunda variável em q . Por linearidade, estendemos estas definições para quaisquer $X_1, \dots, X_l, Y_1, \dots, Y_m \in \mathcal{T}(\mathcal{S})$

$$\begin{aligned} D((X_1 \cdots X_l)_p \parallel q), \quad D(p \parallel (Y_1 \cdots Y_m)_q) \\ \text{e } D((X_1 \cdots X_l)_p \parallel (Y_1 \cdots Y_m)_q). \end{aligned}$$

Agora, nos restringimos para o subconjunto de pontos $(p, p) \in \mathcal{S} \times \mathcal{S}$ e denotamos as funções obtidas em $\mathcal{S}$, novamente para quaisquer $X_1, \dots, X_l, Y_1, \dots, Y_m \in \mathcal{T}(\mathcal{S})$, como

$$\begin{aligned} D[X_1 \cdots X_l \parallel \cdot] : p &\longmapsto D((X_1 \cdots X_l)_p \parallel p) \\ D[\cdot \parallel Y_1 \cdots Y_m] : p &\longmapsto D(p \parallel (Y_1 \cdots Y_m)_p) \\ D[X_1 \cdots X_l \parallel Y_1 \cdots Y_m] : p &\longmapsto D((X_1 \cdots X_l)_p \parallel (Y_1 \cdots Y_m)_p). \end{aligned}$$

Segue da definição (3.14) de D que

$$\begin{aligned} D[\partial_i \parallel \cdot] &= D[\cdot \parallel \partial_i] = 0 \\ D[\partial_i \partial_j \parallel \cdot] &= D[\cdot \parallel \partial_i \partial_j] = -D[\partial_i \parallel \partial_j] (\equiv g_{ij}^{(D)}), \end{aligned}$$

onde a matriz $[g_{ij}^{(D)}]$ é positiva semi-definida. Quando ela é positiva definida em todo ponto de $\mathcal{S}$, chamamos D de *divergência* ou *função contraste* em $\mathcal{S}$.

Dada uma divergência D , uma métrica Riemanniana $g^{(D)} = \langle \cdot, \cdot \rangle^{(D)}$ em $\mathcal{S}$ é definida por $\langle \partial_i, \partial_j \rangle^{(D)} = g_{ij}^{(D)}$, ou equivalentemente por $\langle X, Y \rangle^{(D)} = -D[X \parallel Y]$, para $X, Y \in \mathcal{T}(\mathcal{S})$. A princípio, essa notação pode parecer estranha. Afinal o argumento de D pertence à $\mathcal{S} \times \mathcal{S}$ e acabamos por reescrever derivadas parciais no lugar de pontos da variedade. Todavia, ela ainda vai se mostrar muito conveniente. Esta discussão geral da estrutura em $\mathcal{S}$ vir de uma divergência D é feita por Eguchi.^{36–38}

Esta métrica nos fornece um aproximação de segunda ordem para D dada por

$$D(p \parallel q) = \frac{1}{2} g_{ij}^{(D)}(q) \Delta \theta^i \Delta \theta^j + O(\|\Delta \theta\|^2),$$

onde $\Delta \theta^i \equiv \theta^i(p) - \theta^i(q)$ e $O(\|\Delta \theta\|^2)$ é um termo que vai a zero mais rápido que $\|\Delta \theta\|^2$ quando p tende a q .

Podemos também definir uma conexão afim $\nabla^{(D)}$ via coeficientes dados por

$$\Gamma_{kij}^{(D)} = -D[\partial_i \partial_j \parallel \partial_k],$$

ou equivalentemente por $\langle \nabla_X^{(D)} Y, Z \rangle^{(D)} = -D[XY \parallel Z]$, para $X, Y, Z \in \mathcal{T}(\mathcal{S})$, visto que $\{\Gamma_{kij}^{(D)}\}$ se transformam segundo (2.13). Note que $\nabla^{(D)}$ é necessariamente simétrica ($\Gamma_{kij}^{(D)} = \Gamma_{kji}^{(D)}$), ou seja, livre de torção.

Combinada com a métrica $g^{(D)}$, a conexão $\nabla^{(D)}$ nos fornece uma aproximação de terceira ordem para D dada por

$$D(p \parallel q) = \frac{1}{2} g_{ij}^{(D)}(q) \Delta \theta^i \Delta \theta^j + \frac{1}{6} v_{ijk}^{(D)}(q) \Delta \theta^i \Delta \theta^j \Delta \theta^k + O(\|\Delta \theta\|^3), \quad (3.15)$$

onde $v_{ijk}^{(D)} \equiv D[\partial_i \partial_j \partial_k \parallel \cdot]$, e os coeficiente $\{v_{ijk}^{(D)}\}$ são determinados por $\{g_{ij}^{(D)}\}$ e $\{\Gamma_{kij}^{(D)}\}$ como

$$v_{ijk}^{(D)} = g_{jk,i}^{(D)} + \Gamma_{ijk}^{(D)}. \quad (3.16)$$

De maneira inversa, podemos determinar $g^{(D)}$ e $\nabla^{(D)}$ via expansão (3.15) através da expressão (3.16).

Agora, dizemos que a divergência *dual* a D é identificada como $D^*(p \parallel q) := D(q \parallel p)$. Segue, de forma direta, que

$$g^{(D^*)} = g^{(D)} \quad \text{e} \quad \Gamma_{kij}^{(D^*)} = -D[\partial_k \parallel \partial_i \partial_j].$$

Teorema 5. $\nabla^{(D)}$ e $\nabla^{(D^*)}$ são duais com respeito à $g^{(D)}$.

Prova. A prova segue das definições.

$$\begin{aligned} \partial_k g_{ij}^{(D)} &= -\partial_k D[\partial_i \parallel \partial_j] = -D[\partial_k \partial_i \parallel \partial_j] - D[\partial_i \parallel \partial_k \partial_j] \\ &= -D[\partial_k \partial_i \parallel \partial_j] - D^*[\partial_k \partial_j \parallel \partial_i] \\ &= \Gamma_{jki}^{(D)} + \Gamma_{ikj}^{(D^*)} \end{aligned}$$

□

Podemos estabelecer uma correspondência entre as expressões (3.15) e (3.16) com o caso dual, de forma que

$$\begin{aligned} D^*(q \parallel p) &= D(p \parallel q) \\ &= \frac{1}{2} g_{ij}^{(D)}(p) \Delta \theta^i \Delta \theta^j - \frac{1}{6} v_{ijk}^{(D^*)}(p) \Delta \theta^i \Delta \theta^j \Delta \theta^k + O(\|\Delta \theta\|^3), \end{aligned}$$

onde

$$v_{ijk}^{(D^*)} \equiv D[\cdot \parallel \partial_i \partial_j \partial_k] = g_{jk,i}^{(D)} + \Gamma_{ijk}^{(D^*)}.$$

Portanto, é notável que uma divergência induz uma estrutura dualística livre de torção. Reciprocamente, dada uma tripla (g, ∇, ∇^*) , composta por uma métrica e duas conexões mutuamente duais e simétricas, podemos induzir uma divergência na variedade.

Note que, dada a tripla (g, ∇, ∇^*) e o sistema de coordenadas $[\theta^i]$, definimos

$$D(p \parallel q) \equiv \frac{1}{2}g_{ij}(q)\Delta\theta^i\Delta\theta^j + \frac{1}{6}v_{ijk}(q)\Delta\theta^i\Delta\theta^j\Delta\theta^k,$$

onde $\Delta\theta^i = \theta^i(p) - \theta^i(q)$ e

$$v_{ijk} \equiv g_{jk,i} + \Gamma_{ijk} = \Gamma_{kij} + \Gamma_{jik}^* + \Gamma_{ijk}.$$

Temos então que v_{ijk} é simétrico sobre permutação de todos os seus índices, $D[\partial_i\partial_j \parallel \cdot] = g_{ij}$ e $D[\partial_i\partial_j\partial_k \parallel \cdot] = v_{ijk}$ que imediatamente implica que $(g, \nabla, \nabla^*) = (g^{(D)}, \nabla^{(D)}, \nabla^{(D*)})$. A positividade de D nesse caso vem do fato que g é métrica, ou seja, positiva definida.

Alternativamente, poderíamos ter feito uma escolha dual para D , de forma que

$$D(p \parallel q) \equiv \frac{1}{2}g_{ij}(p)\Delta\theta^i\Delta\theta^j - \frac{1}{6}v_{ijk}^*(p)\Delta\theta^i\Delta\theta^j\Delta\theta^k,$$

onde

$$v_{ijk}^* \equiv g_{jk,i} + \Gamma_{ijk}^* = \Gamma_{kij}^* + \Gamma_{jik} + \Gamma_{ijk}^*.$$

É importantíssimo salientar que toda essa abordagem é local, ou seja, funciona para uma vizinhança aberta do ponto.

3.5.2 As α -divergências

No contexto de $\mathcal{S}$ modelos estatísticos, duas perguntas surgem da discussão da seção anterior: Existe uma divergência globalmente definida em $\mathcal{S}$ que me produziria uma “fábrica” de métricas e conexões duais livre de torção em $\mathcal{S}$? Se sim, como estas métricas se relacionam com a métrica de Fisher e as α -conexões? Nesta seção, queremos responder essas duas perguntas.

Respondendo a primeira pergunta, Matumoto³⁹ provou que em um modelo estatístico toda estrutura dualística livre de torção é induzida de uma divergência globalmente definida. Então, vamos introduzir esta importante classe de divergências.

Seja $h(u)$ função convexa em $\mathbb{R}^+$ tal que $h(1) = 0$, para cada par de distribuições de probabilidade p, q definimos uma h -divergência como

$$D_h(p \parallel q) \equiv \int p(x)h\left(\frac{q(x)}{p(x)}\right)dx. \quad (3.17)$$

A h -divergência deve satisfazer certas propriedades^{40–42}

i) Da desigualdade de Jensen,

$$D_h(p \parallel q) \geq h\left(\int p(x)\frac{q(x)}{p(x)}\right) = h(1) = 0.$$

- ii) D_h é invariante quando $h(u) \rightarrow h(u) + c(u - 1)$, $c \in \mathbb{R}$.
- iii) $D_h^* = D_{h^*}$, onde $f^*(u) = uf(1/u)$.
- iv) D_h é monotônica sobre ação de um mapa qualquer $F : \mathcal{X} \rightarrow \mathcal{Y}$. Se F for estatística suficiente então D é invariante sobre sua ação. Para mais detalhes veja o texto de Amari.²

Quanto a segunda pergunta, vamos assumir h suave. Sendo D_h uma divergência sabemos que ela deve induzir uma métrica $g^{(h)} := g^{(D_h)}$ e uma conexão livre de torção $\nabla^{(h)} := \nabla^{(D_h)}$.

Teorema 6. (Teo. de Chentsov) - Assuma que o conjunto $\{(g_n, \nabla_n)\}_{n=1}^\infty$ é invariante com respeito a estatísticas suficientes, ou seja, para todos n, m , $\mathcal{S} \subset \mathcal{P}_n$ e $F : \mathcal{X}_n \rightarrow \mathcal{X}_m$ estatística invariante para $\mathcal{S}$, métricas induzidas e conexões em $\mathcal{S}$ e $\mathcal{S}_F$ são invariantes. Então, existe $c \in \mathbb{R}^+$ e $\alpha \in \mathbb{R}$, para todo n , tal que g_n coincide com a métrica de Fisher g em $\mathcal{P}_n$, a menos de um fator de escala c , e ∇_n coincide exatamente com $\nabla^{(\alpha)}$ em $\mathcal{P}_n$.

O enunciado do Teorema acima é um ponto de vista diferente daquele encontrado no texto original de Chentsov,⁴³ dele podemos concluir que para modelos estatísticos, a princípio finitos, a métrica de Fisher sob um fator de escala e as α -conexões são caracterizadas pela invariância com respeito a estatísticas suficientes. Assim $g^{(h)} = cg$ e $\nabla^{(h)} = \nabla^{(\alpha)}$, para algum $c > 0$.

Teorema 7. Pelo Teo. 6, as métricas $g^{(h)}$ e as conexões $\nabla^{(h)}$ coincidem com a métrica de Fisher e as α -conexões quando $c = h''(1)$. Fazendo $c = 1$ temos que $\alpha = 3 + 2h'''(1)$.

Prova. Em um modelo estatístico $\mathcal{S}$ os coeficientes da métrica de Fisher são dados por

$$g_{ij}(\theta) = \int p_\theta \partial_i l_\theta \partial_j l_\theta dx = \int \frac{\partial_i p_\theta \partial_j p_\theta}{p_\theta} dx$$

Sabemos que $g_{ij}^{(h)}(\theta) = -D_h[\partial_i \parallel \partial_j](\theta)$. Dado o par de coordenadas (θ^i, θ^i) , de forma que, $p \xrightarrow{\theta^i} p(x, \theta)$ e $q \xrightarrow{\theta^i} q(x, \theta')$. Podemos calcular que

$$\partial'_j \partial_i D_h(p_\theta \parallel q_{\theta'}) = - \int \frac{q_{\theta'}}{p_\theta^2} h'' \partial'_j q_{\theta'} \partial_i p_\theta dx,$$

onde, por simplicidade, vamos deixar sub-entendido o argumento de $h = h(q/p)$ e suas derivadas ao longo dos cálculos. Se $p = q$, então para todo θ

$$-D_h((\partial_i)_\theta \parallel (\partial_j)_\theta) = -D_h[\partial_i \parallel \partial_j](\theta) = g_{ij}^{(h)}(\theta) = \int h''(1) \frac{\partial_i p_\theta \partial_j p_\theta}{p_\theta} dx.$$

Pelo Teo. 6 $g_{ij}^{(h)} = cg_{ij}$. Segue que $c = h''(1)$, pedimos que $c = 1$ para $g^{(h)}$ coincidir exatamente com a métrica de Fisher.

Da mesma forma, sabemos que os coeficientes da conexão livre de torção induzidos por divergência são dados por $(\Gamma_{kij}^{(h)})_\theta = -D_h[\partial_i \partial_j \parallel \partial_k](\theta)$. Os coeficientes das α -conexões podem ser reescritos como

$$\Gamma_{kij}^{(\alpha)} = \int dx \left(\frac{1}{p_\theta} \partial_i \partial_j p_\theta \partial_k p_\theta - \frac{1}{p_\theta^2} \left[\frac{1+\alpha}{2} \right] \partial_i p_\theta \partial_j p_\theta \partial_k p_\theta \right).$$

Podemos calcular que

$$\partial_i \partial'_k \partial_j D(p_\theta \parallel q_{\theta'}) = \int dx \left(\left[2 \frac{q_{\theta'}}{p_\theta^3} h'' + \frac{q_{\theta'}^2}{p_\theta^4} h''' \right] \partial_i p_\theta \partial_j p_\theta \partial'_k q_{\theta'} - \frac{q_{\theta'}}{p_\theta^2} h'' \partial_i \partial_j p_\theta \partial'_k q_{\theta'} \right).$$

Se $p = q$, então para todo θ

$$\begin{aligned} -D_h((\partial_i \partial_j)_\theta \parallel (\partial_k)_\theta) &= -D_h[\partial_i \partial_j \parallel \partial_k](\theta) = (\Gamma_{kij}^{(h)})_\theta \\ &= \int dx \left(\frac{1}{p_\theta} h''(1) \partial_i \partial_j p_\theta \partial_k p_\theta - \frac{1}{p_\theta^2} [2h''(1) + h'''(1)] \partial_i p_\theta \partial_j p_\theta \partial_k p_\theta \right). \end{aligned}$$

Novamente, pelo Teo. 6 $\Gamma_{kij}^{(h)} = \Gamma_{kij}^{(\alpha)}$. Segue que $(1+\alpha)/2 = 2 + h'''(1)$, ou seja, $\alpha = 3 + 2h'''(1)$. $\square$

Vamos então tratar de uma importante família de h -divergências, as chamadas α -divergências $D^{(\alpha)} := D_{h^{(\alpha)}}$ para $\alpha \in \mathbb{R}$, que são definidas pela função contínua em α

$$h^{(\alpha)}(u) = \begin{cases} \frac{4}{1-\alpha^2} [1 - u^{(1+\alpha)/2}], & (\alpha \neq \pm 1) \\ u \log u, & (\alpha = 1) \\ -\log u, & (\alpha = -1) \end{cases}. \quad (3.18)$$

Vale acrescentar, como opinião deste autor, que não é de graça que Amari define essa função $h^{(\alpha)}$ acima. Afinal de contas, tem-se uma série de condições sobre h (h é convexa suave, $h(1) = 0$, $h''(1) = 1$ e $\alpha = 3 + 2h'''(1)$) que nos levaram a recuperar $(g, \nabla^{(\alpha)})$ através das h -divergências de Matumoto. Essas condições vão, de fato, definir uma família de funções dependentes de α , que coincidem exatamente com as funções dadas em $h^{(\alpha)}$ acima. Desta forma determinamos para cada valor de $\alpha \in \mathbb{R}$, ou seja, para cada α -conexão com respeito a métrica de Fisher, qual é a divergência induzida em $\mathcal{S}$.

Temos então que

$$\begin{aligned} D^{(\alpha)}(p \parallel q) &= \frac{4}{1-\alpha^2} \left(1 - \int p(x)^{\frac{1-\alpha}{2}} q(x)^{\frac{1+\alpha}{2}} dx \right), \quad (\alpha \neq \pm 1) \\ D^{(-1)}(p \parallel q) &= D^{(1)}(p \parallel q) = \int p(x) \log \frac{p(x)}{q(x)} dx. \end{aligned} \quad (3.19)$$

Claramente vamos ter que $(g^{(h^{(\alpha)})}, \nabla^{(h^{(\alpha)})}) = (g, \nabla^{(\alpha)})$. Também temos que, $D^{(\alpha)}(p \parallel q) = D^{(-\alpha)}(q \parallel p)$, refletindo a dualidade entre $\nabla^{(\alpha)}$ e $\nabla^{(-\alpha)}$. Em particular temos que

$D^{(0)}(p \parallel q)$ é simétrico, refletindo que para $\alpha = 0$ a conexão é Riemanniana com respeito à métrica Fisher, e além disto $\sqrt{D^{(0)}(p \parallel q)}$ satisfaz os axiomas de distância, de modo que

$$D^{(0)}(p \parallel q) = 2 \int \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2 dx. \quad (3.20)$$

$\sqrt{D^{(0)}(p \parallel q)}$ ou $D^{(0)}(p \parallel q)$, são chamadas de *distância de Hellinger*.

As α -divergências são muito próximas as α -entropias de Rényi,⁴⁴ distância de Chernoff⁴⁵ e entropia de Tsallis.⁴⁶⁻⁴⁸ Além do mais, as ± 1 -divergências são conhecidas por vários nomes distintos pela literatura, a dizer, divergências de Kullback, informação de Kullback-Leibler, entropia relativa ou simplesmente divergência.

Apenas por tratar geometricamente conjuntos de distribuições de probabilidade, conseguimos recuperar de maneira sistemática e geral resultados já conhecidos pela literatura.

3.6 Geometria da informação para sistemas quânticos

3.6.1 Espaço de estados

De forma sucinta, a teoria quântica descreve a física de um sistema através de operadores lineares especiais em um espaço de matrizes complexas. Seja então o espaço de Hilbert $\mathcal{H} \simeq \mathbb{C}^k$ de dimensão finita $\dim \mathcal{H} = k$. Utilizando brevemente a notação de Dirac, sendo o produto interno de $|\phi\rangle, |\psi\rangle \in \mathcal{H}$ denotado por $\langle \phi | \psi \rangle$, definimos o hermitiano conjugado $A^\dagger$ de um operador A pela propriedade $\langle \phi | A \psi \rangle = \langle A^\dagger \phi | \psi \rangle$. Então, apresentamos o espaço de estados quântico como o conjunto convexo de operadores ρ em $\mathcal{H}$ hermitianos, positivos semi-definidos, de traço unitário, ou seja,

$$\rho = \rho^\dagger, \quad \rho \geq 0 \quad \text{e} \quad \text{Tr} \rho = 1.$$

Elementos deste conjunto representam estados quântico do nosso sistema. Dado estes operadores, chamamos um estado de *puro* quando ρ é matriz com $\text{rank}(\rho) = 1$, representada como $\rho = |\psi\rangle \langle \psi|$ para algum $|\psi\rangle \in \mathcal{H}$, caso contrário chamamos o estado de *misto*. Uma definição equivalente para o caso puro é dizer que $\rho = \rho^2$.

Uma *medida* será expressa matematicamente por $\Pi = \{\pi_i\}_{i=1,\dots,l}$ conjunto de operadores em $\mathcal{H}$ que satisfaz

$$\pi_i = \pi_i^\dagger \geq 0 \quad \text{e} \quad \sum_i \pi_i = \mathbf{1},$$

onde $\mathbf{1}$ é o operador identidade em $\mathcal{H}$.

Dado que o estado do sistema é representado por ρ , quando realizamos uma medida Π sobre um sistema quântico produzimos um elemento do conjunto finito de resultados $\{e_1, \dots, e_l\}$ possíveis, com a probabilidade de e_i ser observado descrita por

$$p_i = p(e_i) = \text{Tr}(\rho \pi_i).$$

Note a condição de positividade semi-definida de ρ e π_i implica que $p_i \geq 0$ e o fato da soma dos π_i comporem a identidade implica que $\sum_i p_i = 1$.

De maneira geral, uma medida que toma resultados em um espaço mensurável $(\mathcal{X}, \mathcal{B})$, onde $\mathcal{B}$ é σ -álgebra consistindo de subconjuntos de $\mathcal{X}$, é representada por um mapa π que leva um conjunto mensurável $B \in \mathcal{B}$ em $\pi(B) = \pi(B)^\dagger \geq 0$ operador hermitiano em $\mathcal{H}$, satisfazendo $\pi(\mathcal{X}) = \mathbf{1}$ e para um família mutualmente disjunta $\{B_1, B_2, \dots\} \in \mathcal{B}$

$$\pi\left(\bigcup_i B_i\right) = \sum_i \pi(B_i),$$

ou seja, π é completamente aditiva.

Tal mapa é denotado como sendo um *POVM*.[†] A probabilidade de uma medida cair dentro de um conjunto B é dada por $p(B) = \text{Tr}[\rho\pi(B)]$. Fica claro que quando $\mathcal{X} = \{e_1, \dots, e_l\}$ e $\mathcal{B}$ é o conjunto das parte de $\mathcal{X}$ temos $\pi_i = \pi(\{e_i\})$.

Dizemos que Π é uma *medida simples* ou *medida projetiva* quando a medida satisfaz as condições adicionais

$$\pi_i^2 = \pi_i \quad \text{e} \quad \pi_i \pi_j = 0 \quad \text{para } i \neq j,$$

ou seja, quando a medida consiste de projeções ortogonais em subespaços mutualmente ortogonais de $\mathcal{H}$. Dada uma medida simples, vamos rotular seus possíveis resultados $\{e_1, e_2, \dots\}$ por uma sequência de números reais distintos $\{a_1, a_2, \dots\}$. Vamos chamar o par $(\Pi, \{a_i\})$ de *observável*, sendo Π medida simples, representado pelo operador hermitiano

$$A = \sum_i a_i \pi_i.$$

Fica claro que os $\{a_i\}$ são os autovalores de A e $\{\pi_i\}$ as projeções ortogonais nos autoespaços correspondentes, além disso, temos bijeção entre um observável A e $(\Pi, \{a_i\})$ com Π medida simples. Um cálculo direto nos faz ver que o valor esperado e a variância de um observável A atuando no sistema descrito pelo estado ρ são dados por

$$\begin{aligned} E_\rho[A] &= \sum_i a_i \text{Tr}(\rho \pi_i) = \sum_i a_i p_i = \text{Tr}(\rho A) \quad \text{e} \\ \text{Var}_\rho[A] &= \sum_i \text{Tr}(\rho \pi_i) (a_i - E_\rho[A])^2 = \text{Tr}(\rho A^2) - \text{Tr}(\rho A)^2. \end{aligned}$$

Vamos agora apresentar notação acerca de alguns conjuntos que serão recorrentes. Denotamos por

$$\mathcal{A} \equiv \{A \mid A = A^\dagger\}$$

o conjunto de todos operadores hermitianos em $\mathcal{H}$ e

$$\mathcal{A}_1 \equiv \{A \mid A \in \mathcal{A} \quad \text{e} \quad \text{Tr} A = 1\}$$

[†]Sigla vinda do inglês *positive-operator valued measure* que será utilizada para nos referirmos à medidas generalizadas.

sua restrição para operadores de traço unitário.

O conjunto total de operadores densidade

$$\bar{\mathcal{S}} \equiv \{\rho \mid \rho = \rho^\dagger \geq 0 \quad \text{e} \quad \text{Tr}\rho = 1\},$$

é subconjunto convexo de $\mathcal{A}_1$. Este conjunto pode ser particionado em $\bar{\mathcal{S}} = \bigcup_{r=1}^k \mathcal{S}_r$, onde $\mathcal{S}_r \equiv \{\rho \in \bar{\mathcal{S}} \mid \text{rank}(\rho) = r\}$. Pontos de $\mathcal{S}_1$ correspondem a estados puros e são chamados de pontos extremais, ou seja, pontos que não podem ser representados por combinação convexa de outros pontos. $\mathcal{S}_k$ é o conjunto de operadores densidade positivos definidos e corresponde ao interior do convexo.⁴⁹

O conjunto de todos os operadores unitários em $\mathcal{H}$ é denotado por

$$\mathcal{U} = \{U \mid U^{-1} = U^\dagger\}.$$

Esse conjunto forma um grupo de Lie, portanto já é uma variedade em si, agindo em $\bar{\mathcal{S}}$ por

$$\begin{aligned} \mathcal{U} \times \bar{\mathcal{S}} &\rightarrow \bar{\mathcal{S}} \\ (U, \rho) &\longmapsto U\rho U^\dagger, \end{aligned}$$

de forma que todo $\mathcal{S}_r$ é fechado por esta ação. Note, um estado ρ , uma medida simples Π e seu respectivo observável A são transformados em

$$\tilde{\rho} = U\rho U^\dagger, \quad \tilde{\pi}_i = U\pi_i U^\dagger \quad \text{e} \quad \tilde{A} = UAU^\dagger.$$

Pela unitariedade de U vemos que as distribuições de probabilidade e os valores esperados são invariantes sob a ação do grupo unitário,

$$\text{Tr}(\tilde{\rho}\tilde{\pi}_i) = \text{Tr}(\rho\pi_i) \quad \text{e} \quad \text{Tr}(\tilde{A}\tilde{\rho}) = \text{Tr}(A\rho).$$

A geometria do espaço de estados puros $\mathcal{S}_1$ pode ser identificada com a do espaço projetivo complexo $\mathbb{P}_{\mathbb{C}}^{k-1}$ que possui métrica única invariante sob ação do grupo unitário chamada *métrica de Fubini-Study*, vide apêndice D. Contudo, aqui estamos interessados na geometria de $\mathcal{S}_k$, que agora será denotado apenas por

$$\mathcal{S} \equiv \{\rho \mid \rho = \rho^\dagger > 0 \quad \text{e} \quad \text{Tr}\rho = 1\}.$$

Este espaço é subconjunto de $\mathcal{A}_1$ e naturalmente dotado de estrutura de variedade real (não complexa, pois pode ter dimensão ímpar)⁵⁰ de dimensão $\dim \mathcal{S} = k^2 - 1$.

3.6.2 Caracterização do espaço tangente

Queremos caracterizar como são os vetores tangentes a um ponto $\rho \in \mathcal{S}$. Primeiramente vamos enunciar teoremas que vão nos auxiliar a caracterizar o espaço tangente nesse

caso, a busca das provas destes resultados fica a critério do leitor afeito a matemática e podem ser encontradas em um bom livro de geometria diferencial.^{51,52}

Dado o mapa suave $\alpha : M \rightarrow N$, com M e N variedades diferenciáveis de dimensão m e n respectivamente. Em $p \in M$ induzimos naturalmente um mapa linear $T_p\alpha : T_p(M) \rightarrow T_{\alpha(p)}(N)$ tal que, dado $f \in \mathcal{F}_{\alpha(p)}(N)$ e $V_p \in T_p(M)$,

$$T_p\alpha(V_p)(f) \equiv V_p(f \circ \alpha)$$

satisfazendo as seguintes

- i) Dado que $M \xrightarrow{\alpha} N \xrightarrow{\delta} P$ e $p \in M$, então $T_p(\delta \circ \alpha) = T_{\alpha(p)}\delta \circ T_p\alpha$.
- ii) $T_p Id_M = Id_{T_p(M)}$.
- iii) Se α é difeomorfismo, então $T_p\alpha^{-1} = (T_p\alpha)^{-1}$.

Onde Id é denota o mapa identidade.

Dado $\epsilon \in \mathbb{R}$, seja a curva $\gamma : (-\epsilon, \epsilon) \rightarrow M$ tal que $\gamma(0) = p \in M$ e $\dot{\gamma}_p(0) = V_p \in T_p(M)$. Dado o mapa $\alpha : M \rightarrow N$ e $f \in \mathcal{F}_{\alpha(p)}(N)$, onde identificamos $\beta(t) := \alpha(\gamma(t))$, da definição (2.6) temos que

$$\begin{aligned} \dot{\beta}_{\alpha(p)}(0)(f) &= \left. \frac{d}{dt} \right|_{t=0} f(\beta(t)) = \left. \frac{d}{dt} \right|_{t=0} f(\alpha(\gamma(t))) \\ &= \dot{\gamma}_p(0)(f \circ \alpha) = V_p(f \circ \alpha) = T_p\alpha(V_p)(f). \end{aligned}$$

Portanto, como a relação acima vale para qualquer f ,

$$T_p\alpha(V_p) = \left. \frac{d}{dt} \right|_{t=0} \alpha(\gamma(t)). \quad (3.21)$$

Dado o mapa suave $\alpha : M \rightarrow N$. Definimos que $p \in M$ é *ponto regular* de α se $\text{rank}(T_p\alpha) = \dim N$, caso contrário p é *ponto crítico*. Ainda, $q \in N$ é dito *valor regular* de α se todo $p \in \alpha^{-1}(q)$ é ponto regular, caso contrário q é *valor crítico*.

Teorema 8. Se q é valor regular de $\alpha : M \rightarrow N$ então $\alpha^{-1}(q) = S$ é *subvariedade* de M com $\text{codim}_M S = \dim M - \dim S = \dim N$.

Teorema 9. Seja q valor regular de $\alpha : M \rightarrow N$ e $s \in S$, onde $S = \alpha^{-1}(q)$ é subvariedade de M . Então $T_s(S) \subset \text{Ker}(T_s\alpha)$. Onde Ker denota o núcleo da transformação.

Aplicando esses resultados ao nosso contexto, seja o mapa $\alpha : \mathcal{A} \rightarrow \mathbb{R}$, de forma que $\alpha(A) = \text{Tr}(A)$, onde $\mathcal{A} \subset M_n(\mathbb{C}) \simeq \mathbb{C}^{k^2}$. Dado que $1 \in \mathbb{R}$ é valor regular de α , pelo Teo. 8, $\mathcal{A}_1 = \alpha^{-1}(1)$ é subvariedade de $\mathcal{A}$ com dimensão $\dim \mathcal{A}_1 = \dim \mathcal{A} - \dim \mathbb{R} = k^2 - 1$.

Seja um vetor tangente $B_A \in T_A(\mathcal{A}_1)$ ao ponto $A \in \mathcal{A}_1$ definido pela curva $\gamma(t) = A + tB_A$ em $\mathcal{A}$, com $t \in (-\epsilon, \epsilon) \subset \mathbb{R}$. Pelo Teo. 9, $T_A(\mathcal{A}_1) \subset \text{Ker}(T_A\alpha)$, então pelo resultado (3.21) temos que

$$T_A\alpha(B_A) = \left. \frac{d}{dt} \right|_{t=0} \text{Tr}(A + tB_A) = \text{Tr}(B_A) = 0.$$

Portanto, vetores tangentes em $\mathcal{A}_1$ são caracterizados por matrizes de traço nulo, ou seja, $T_A(\mathcal{A}_1) \simeq \mathcal{A}_0$. Como $\mathcal{S}$ é conjunto aberto de $\mathcal{A}_1$ então também é variedade aberta $(k^2 - 1)$ -dimensional, como já foi mencionado, com $T_\rho(\mathcal{S})$ em cada $\rho \in \mathcal{S}$ identificado com $\mathcal{A}_0 = \{A \mid A \in \mathcal{A} \text{ e } \text{Tr}A = 0\}$. Note que, usando raciocínios análogos ao acima, pode-se provar que vetores tangentes a elementos dos grupos de Lie $O(n)$ e $\mathcal{U}$ são caracterizados, respectivamente, por matrizes antissimétricas e anti-hermitianas.

Agora, vamos mostrar que o espaço $T_\rho(\mathcal{S})$ deve se quebrar em duas direções, uma comutativa e outra não-comutativa. Dado uma medida simples $\Pi = \{\pi_i\}_{i=1,\dots,l}$, definimos

$$\begin{aligned} \mathcal{S}_\Pi &\equiv \mathcal{S} \cap \left\{ \sum_{i=1}^l a_i \pi_i \mid (a_1, \dots, a_l) \in \mathbb{R}^l \right\} \\ &= \left\{ \sum_{i=1}^l \frac{p_i}{\text{Tr}(\pi_i)} \pi_i \mid p = (p_1, \dots, p_l) \in \mathcal{P}_l \right\}, \end{aligned}$$

onde $\mathcal{P}_l$ denota $\mathcal{P}(\{1, \dots, l\})$, totalidade de distribuições de probabilidade em $\{1, \dots, l\}$. Se $\rho, \sigma \in \mathcal{S}_\Pi$ então $\rho\sigma = \sigma\rho$. Temos que $\mathcal{S}_\Pi$ é subvariedade $(l - 1)$ -dimensional comutativa de $\mathcal{S}$, qualquer modelo comutativo deve ser subvariedade de $\mathcal{S}_\Pi$ para alguma medida simples Π . Medidas simples diagonalizam $\mathcal{S}$.

Ainda, se todos $\pi_i \in \Pi$ tem $\text{rank}(\pi_i) = 1$ para todo i ,

$$\mathcal{S} = \{U\rho U^\dagger \mid \rho \in \mathcal{S}_\pi \text{ e } U \in \mathcal{U}\}.$$

Dado o fato acima, vamos fazer um parêntesis na discussão para apresentar um cálculo totalmente informal, mas muito ilustrativo, que vai nos dar a intuição do fato que o espaço tangente se quebrará em duas direções. Seja $\rho \in \mathcal{S}$, existe $U \in \mathcal{U}$ tal que $U^\dagger \rho U = \text{diag} \left(\frac{p_1}{\text{Tr}(\pi_1)}, \frac{p_2}{\text{Tr}(\pi_2)}, \dots \right) \equiv \rho_D \in \mathcal{S}_\Pi$ para alguma medida simples Π . Indicando derivação por d , pela unitariedade de U segue que $U^\dagger dU = -U dU^\dagger$, então

$$\begin{aligned} d\rho &= d(U\rho_D U^\dagger) = U d\rho_D U^\dagger - \rho(U^\dagger dU) + (U^\dagger dU)\rho \\ &= d^C \rho_D + [\Omega, \rho], \end{aligned}$$

onde $\Omega := U^\dagger dU = -U dU^\dagger$ e $d^C \rho_D = \text{diag} \left(\frac{dp_1}{\text{Tr}(\pi_1)}, \frac{dp_2}{\text{Tr}(\pi_2)}, \dots \right)$. Vemos que derivações de ρ se quebram em uma parte comutativa e um comutador de ρ com um operador anti-hermitiano.

Formalmente, seja U_t uma curva em $\mathcal{U}$ satisfazendo $U_0 = \mathbf{1}$ e $\dot{U}_t|_{t=0} = \Omega$. Definindo a curva $\rho_t \equiv U_t \rho U_t^\dagger$ passando por $\rho \in \mathcal{S}$ em $t = 0$, temos que um vetor tangente em ρ deve ser dado por

$$\left. \frac{d}{dt} \right|_{t=0} U_t \rho U_t^\dagger = [\Omega, \rho].$$

Assim, definimos o subespaço $T_\rho^u \subset T_\rho(\mathcal{S})$ consistindo de todas as direções tangentes não-comutativas correspondentes a ação de $\mathcal{U}$ como

$$T_\rho^u = \{X_\rho \in T_\rho(\mathcal{S}) \mid \exists \Omega = -\Omega^\dagger \text{ tal que } X_\rho = [\Omega, \rho]\}.$$

Note que se $\Omega = -iH$, com H hermitiano, a expressão que caracteriza o vetor tangente a curva no ponto ρ é exatamente a dinâmica de von Neumann.

Todos os vetores tangentes em ρ que não estão nas direções de T_ρ^u deverão estar na direção comutativa. Definimos o subespaço $T_\rho^c \subset T_\rho(\mathcal{S})$ consistindo de todas as direções tangentes comutativas como

$$T_\rho^c = \{X_\rho \in T_\rho(\mathcal{S}) \mid [\rho, \dot{\rho}_t|_{t=0}] = 0\}.$$

Em outras palavras, T_ρ^c é a totalidade das derivadas $d\rho_t/dt|_{t=0}$ de curvas ρ_t satisfazendo $\rho_0 = \rho$ e $[\rho, \rho_t] = 0$ para todo t . Para uma medida simples Π que diagonaliza ρ , é evidente que $T_\rho(\mathcal{S}_\Pi) \subset T_\rho^c$, se ainda ρ tiver k autovalores distintos então $T_\rho(\mathcal{S}_\Pi) = T_\rho^c$.

Definindo o mapa linear $\Lambda : \mathcal{A}_0 \rightarrow \mathcal{A}_0$ por $\Lambda(A) = i[\rho, A]$, vemos que Λ é antissimétrico com respeito ao produto interno de Hilbert-Schmidt $\ll A, B \gg_{HS} \equiv \text{Tr}(AB)$, com $A, B \in \mathcal{A}_0$, e que T_ρ^c e T_ρ^u são respectivamente o núcleo e a imagem de Λ . Portanto, temos que $T_\rho(\mathcal{S})$ é decomposto na soma direta desses dois subespaços, que vão ser complementos ortogonais um do outro,

$$T_\rho(\mathcal{S}) = T_\rho^c \oplus T_\rho^u.$$

3.6.3 Divergência e métricas quânticas

Queremos então introduzir uma estrutura dualística em $\mathcal{S}$ que será o análogo quântico da métrica de Fisher e $\pm\alpha$ -conexões. Sabemos que toda estrutura dualística livre de torção é induzida de uma divergência. Por isso, antes devemos definir uma divergência quântica similar à (3.17) e aplicarmos ao caso específico da estrutura $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ induzida de funções convexas suaves $h : \mathbb{R}^+ \rightarrow \mathbb{R}$ satisfazendo

$$h(1) = 0, \quad h''(1) = 1 \quad \text{e} \quad \alpha = 3 + 2h'''(1), \quad (3.22)$$

que definem as α -divergências $D^{(\alpha)}$, como visto na Sec. 3.5.

Seja $\mathcal{B}$ o conjunto de todos os operadores em $\mathcal{H}$. Dados $\sigma, \rho \in \mathcal{S}$, o *operador relativo modular* $\Delta = \Delta_{\sigma, \rho} : \mathcal{B} \rightarrow \mathcal{B}$ é definido como

$$\Delta(A) = \sigma A \rho^{-1}, \quad (3.23)$$

ou equivalentemente por

$$\ll \Delta(A), B \gg_{\rho}^{-} = \ll A, B \gg_{\sigma}^{+} \quad \text{com } A, B \in \mathcal{B},$$

onde $\ll \cdot, \cdot \gg_{\rho}^{-}$ e $\ll \cdot, \cdot \gg_{\sigma}^{+}$ são produtos internos em $\mathcal{B}$ onde

$$\ll A, B \gg_{\rho}^{-} = \text{Tr}(\rho A^{\dagger} B) \quad \text{e} \quad \ll A, B \gg_{\sigma}^{+} = \text{Tr}(\sigma B A^{\dagger}).$$

Desde que Δ é operador hermitiano e positivo definido no espaço de Hilbert $(\mathcal{B}, \ll \cdot, \cdot \gg_{\rho}^{-})$, nesse caso, para uma função $h : \mathbb{R}^{+} \rightarrow \mathbb{R}$ vale a seguinte

$$\Delta(A) = \lambda A \rightarrow h(\Delta)(A) = h(\lambda)A, \quad (3.24)$$

onde $\lambda \in \mathbb{R}^{+}$ e $A \in \mathcal{B}$.

Desta forma, definimos a *h-divergência quântica* D_h como

$$D_h(\rho \parallel \sigma) \equiv \ll \mathbf{1}, h(\Delta)(\mathbf{1}) \gg_{\rho}^{-} = \text{Tr}[\rho h(\Delta)(\mathbf{1})]. \quad (3.25)$$

Note, usando a decomposição espectral

$$\rho = \sum_i a_i \mu_i \quad \text{e} \quad \sigma = \sum_j b_j \nu_j,$$

onde $\{\mu_i\}$ e $\{\nu_j\}$ são medidas simples e $\{a_i\}$ e $\{b_j\}$ autovalores, temos que

$$\Delta(\nu_j \mu_i) = \frac{b_j}{a_i} \nu_j \mu_i \quad \text{e} \quad h(\Delta)(\nu_j \mu_i) = h\left(\frac{b_j}{a_i}\right) \nu_j \mu_i.$$

Dadas as medidas $M = \{M_{ij} = \mu_i \nu_j \mu_i\}$ e $N = \{N_{ij} = \nu_j \mu_i \nu_j\}$ identificamos o conjunto de probabilidades $p = \{p_{ij}\}$ e $q = \{q_{ij}\}$ como

$$p_{ij} = \text{Tr}(\rho M_{ij}) = a_i \text{Tr}(\mu_i \nu_j) \quad \text{e} \quad q_{ij} = \text{Tr}(\sigma N_{ij}) = b_j \text{Tr}(\mu_i \nu_j).$$

O que nos leva a

$$D_h(\rho \parallel \sigma) = \sum_{i,j} a_i h\left(\frac{b_j}{a_i}\right) \text{Tr}(\mu_i \nu_j) = \sum_{i,j} p_{ij} h\left(\frac{q_{ij}}{p_{ij}}\right).$$

A equação acima indica que $D_h(\rho \parallel \sigma)$ irá coincidir exatamente com a *h-divergência clássica* de Matumoto na Eq.(3.17), quando h é função convexa satisfazendo $h(0) = 1$, assim D_h se torna uma divergência em $\mathcal{S}$ no sentido descrito na Sec. 3.5.

Na notação da Sec. 3.5, assumindo que h é convexa suave satisfazendo as condições (3.22), a estrutura induzida $(g^{(h)}, \nabla^{(h)}, \nabla^{(h*)})$ é um análogo quântico formal da tripla $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$. Além disso, nos restringimos à $\mathcal{S}_{\Pi}$ faz D_h ser idêntica as *h-divergências clássicas* em $\mathcal{P}_l$ sobre a identificação natural $\mathcal{S}_{\Pi} \simeq \mathcal{P}_l$. Outro fato notável é que D_h é invariante sobre a ação de $\mathcal{U}$, ou seja, $D_h(U\rho U^{\dagger} \parallel U\sigma U^{\dagger}) = D_h(\rho \parallel \sigma)$.

Então, com h convexa suave satisfazendo as condições (3.22), podemos enunciar que a estrutura dualística quântica satisfaz as seguintes

- i) Para uma medida simples Π a restrição da estrutura dualística em $\mathcal{S}_\Pi$ coincide, para algum α , com $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ em $\mathcal{P}_l$ sobre a identificação natural $\mathcal{S}_\Pi \simeq \mathcal{P}_l$.
- ii) $(g, \nabla^{(\alpha)}, \nabla^{(-\alpha)})$ são invariantes sob ação de $\mathcal{U}$. Dado $U \in \mathcal{U}$ e X, Y campos vetoriais suaves em $\mathcal{S}$

$$\begin{aligned}\langle UXU^\dagger, UYU^\dagger \rangle &= \langle X, Y \rangle \\ \nabla_{UXU^\dagger} UYU^\dagger &= U(\nabla_X Y)U^\dagger \\ \nabla_{UXU^\dagger}^* UYU^\dagger &= U(\nabla_X^* Y)U^\dagger,\end{aligned}$$

onde $UXU^\dagger$ é campo vetorial tal que $(UXU^\dagger)_{U\rho U^\dagger} = UX_\rho U^\dagger$.

Agora vamos fazer as funções convexas suaves h assumirem a forma $h^{(\alpha)}$ vista em (3.18). Sobre o operador relativo modular, serão úteis as seguintes

$$(\Delta)^r(A) = \sigma^r A \rho^{-r},$$

para todo $r \in \mathbb{R}$, e

$$(\log \Delta)(A) = (\log \sigma)A - A(\log \rho).$$

Assim obtemos as α -divergências quânticas

$$\begin{aligned}D^{(\alpha)}(\rho \parallel \sigma) &= \frac{4}{1 - \alpha^2} \{1 - \text{Tr}(\rho^{\frac{1-\alpha}{2}} \sigma^{\frac{1+\alpha}{2}})\} \quad (\alpha \neq \pm 1) \\ D^{(-1)}(\rho \parallel \sigma) &= D^{(1)}(\sigma \parallel \rho) = \text{Tr}[\rho(\log \rho - \log \sigma)],\end{aligned}$$

correspondentes às divergências clássicas (3.19), informalmente, apenas trocando a integral pelo traço. Este formato das α -divergências quânticas foi apresentado primeiramente por Hasegawa.⁵³

Dado que nossa divergência quântica produz, de fato, um análogo quântico formal da tripla Fisher e $\pm\alpha$ -conexões, temos que uma tradução da discussão vista na Sec. 3.3 sobre a origem afim das α -conexões se aplica aqui, de onde podemos extrair facilmente uma expressão para a métrica.

Então, parametrizando os elementos de $\mathcal{S}$ como ρ_θ por um sistema de coordenadas $[\theta^i]$, para $\theta \in \Theta \subseteq \mathbb{R}^n$, e dado a família de mergulhos para $\alpha \in \mathbb{R}$

$$\begin{aligned}L^{(\alpha)}(\sigma) &\equiv \begin{cases} \frac{2}{1 - \alpha} \sigma^{(\frac{1-\alpha}{2})}, & (\alpha \neq 1) \\ \log(\sigma), & (\alpha = 1) \end{cases} \\ \ell_{\rho_\theta}^{(\alpha)} &\equiv L^{(\alpha)}(\rho_\theta),\end{aligned}\tag{3.26}$$

identificamos $\mathcal{S}$ com o conjunto $\{\ell_{\rho_\theta}^{(\alpha)} \mid \rho_\theta \in \mathcal{S}\}$. Denotaremos $\ell_{\rho_\theta}^{(\alpha)}$ apenas por $\ell^{(\alpha)}$.

Um vetor tangente $X \in T_\theta(\mathcal{S})$ é então representado pelo resultado da operação de X sobre $\rho \mapsto \ell^{(\alpha)}$ e deve valer $X^{(\alpha)} = X\ell^{(\alpha)}$. Já as componentes da métrica $g^{(\alpha)}$ e da conexão $\nabla^{(\alpha)}$ são dadas por

$$\begin{aligned} g_{ij}^{(\alpha)} &= \text{Tr}[\partial_i \ell^{(\alpha)} \partial_j \ell^{(-\alpha)}], \\ \Gamma_{kij}^{(\alpha)} &= \text{Tr}[\partial_i \partial_j \ell^{(\alpha)} \partial_k \ell^{(-\alpha)}], \end{aligned}$$

correspondentes das relações (3.10) e (3.11), com $\partial_i = \partial/\partial\theta^i$.

Caracterizamos $T_\theta^{(\alpha)u}$ como o conjunto de todas as derivadas $d\ell_t^{(\alpha)}/dt|_{t=0} = [\Omega, \ell^{(\alpha)}]$ correspondentes a ação de $\mathcal{U}$, com Ω anti-hermitiano, e $T_\theta^{(\alpha)c}$ como o conjunto de todas as derivadas $d\ell_t^{(\alpha)}/dt|_{t=0}$ de curvas $\ell_t^{(\alpha)}$ tal que $\ell_0^{(\alpha)} = \ell^{(\alpha)}$ e $[\ell^{(\alpha)}, \dot{\ell}_t^{(\alpha)}|_{t=0}] = 0$. Olhando apenas para $T_\theta^{(\alpha)u}$, sendo o vetor tangente $\partial_i \ell^{(\alpha)} = [\Omega_i, \ell^{(\alpha)}]$, obtemos

$$g_{ij}^{(\alpha)u}(\theta) = \begin{cases} \frac{4}{1-\alpha^2} \text{Tr}([\Omega_i, \rho_\theta^{\frac{1+\alpha}{2}}][\Omega_j, \rho_\theta^{\frac{1-\alpha}{2}}]), & (\alpha \neq 1) \\ \text{Tr}([\Omega_i, \log \rho_\theta][\Omega_j, \rho_\theta]) & (\alpha = \pm 1). \end{cases} \quad (3.27)$$

Olhando agora para $T_\theta^{(\alpha)c}$, sendo o vetor tangente comutativo $\partial_i \ell^{(\alpha)} \equiv \partial_i^C \ell^{(\alpha)}$, obtemos

$$g_{ij}^c(\theta) = \text{Tr} \left[\frac{\partial_i^C \rho_\theta \partial_j^C \rho_\theta}{\rho_\theta} \right] \quad \text{para todo } \alpha \in \mathbb{R}, \quad (3.28)$$

recuperando a expressão de Fisher clássica em $\mathcal{P}_k$ como esperado. Note que nesse caso comutativo, dado uma medida simples Π que diagonaliza ρ_θ composta de projeções em uma base ortonormal de $\mathcal{H}$, com ρ_θ tendo k autovalores distintos, temos que $\rho_\theta = \text{diag}(p_1(\theta), \dots, p_k(\theta))$; assim a métrica acima se reescreve como

$$g_{ij}^c(\theta) = \sum_{r=1}^k \frac{\partial_i p_r(\theta) \partial_j p_r(\theta)}{p_r}. \quad (3.29)$$

Além disso, seja $\partial_i \ell^{(\alpha)} = [\Omega_i, \ell^{(\alpha)}] \in T_\theta^{(\alpha)u}$ e $V^{C(\alpha)} \in T_\theta^{(\alpha)c}$ satisfazendo por definição $[\ell^{(\alpha)}, V^{C(\alpha)}] = 0$, então

$$\langle \partial_i \ell^{(\alpha)}, V^{C(\alpha)} \rangle_\theta = \text{Tr}([\Omega_i, \ell^{(\alpha)}] V^{C(\alpha)}) = \text{Tr}(\Omega_i [\ell^{(\alpha)}, V^{C(\alpha)}]) = 0.$$

Portanto $T_\theta^{(\alpha)u}$ e $T_\theta^{(\alpha)c}$ são direções ortogonais de $T_\theta(\mathcal{S})$ para todo α . Em particular, podemos enunciar que a estrutura dualística quântica satisfaz também

iii) T_ρ^c e T_ρ^u são ortogonais com respeito a g em todo ponto $\rho \in \mathcal{S}$.

Escrevendo a métrica como a soma direta

$$g_{ij}^{(\alpha)} = g_{ij}^c \oplus g_{ij}^{(\alpha)u}, \quad (3.30)$$

evidenciamos que a ela se quebra em uma parte correspondente as direções comutativa $T_\theta^{(\alpha)c}$ de $\mathcal{S}$, dada precisamente pela métrica de Fisher, e uma parte correspondente as direções

não-comutativas da álgebra de matrizes $T_\theta^{(\alpha)u}$, composta de uma família de métricas vindas das coerências dos estados quânticos. Temos então uma expressão da métrica para cada α -mergulho.

Vale destacar que Hasegawa⁵³ deriva estas expressões para métrica sem lançar mão de modelos paramétricos de $\mathcal{S}$, porém esta derivação é bem mais complexa.

3.6.4 Discussão sobre a contratividade das α -divergências

Nos vimos na Sec. 3.5.2 que as h -divergências clássicas são monotônicas sobre ação de um mapa qualquer $F : \mathcal{X} \rightarrow \mathcal{Y}$. Aqui vamos apresentar uma discussão deste fato estendido ao caso quântico.

Quanticamente é desejado contratividade métrica sob ação de mapa CPTP.^{†14, 49, 54, 55} São mapas gerais, não necessariamente dinâmicos, que representam mais das vezes efeitos de decoerência pela ação do ambiente, que diminuiria a distinguibilidade (distância) entre dois estados.

Petz⁵⁶ afirma que se h é *operador convexa* então a h -divergência quântica D_h é contrativa sob ação de mapa CPTP $\epsilon(\cdot)$, de forma que $D_h(\epsilon(\rho) \parallel \epsilon(\sigma)) \leq D(\rho \parallel \sigma)$. Hasegawa⁵³ afirma que as funções $h^{(\alpha)}$ são operadores convexas para $\alpha \in [-3, 3]$, indicando a prova do fato ao texto de Bendat & Sherman⁵⁷ sem apontar qual dos teoremas usou. Acreditamos que ele deva ter usado o Teorema 3.4 para isso. É importante frisar que se a divergência é contrativa então a métrica associada a ela também o será, dada a relação entre elas vista na Sec. 3.4. Contudo, um fato muito fácil de provar a partir dos textos de Hasegawa e Petz^{53, 56} que para $\alpha \in [-1, 1]$ as $h^{(\alpha)}$ são *operadores função conjuntamente convexas*, bastando usar o Corolário de Ando do Teorema da concavidade de Lieb.^{58, 59}

No caso $\alpha \in [-1, 1]$, a menos de um fator constante, recuperamos a informação de Wigner-Yanase-Dyson-Lieb da relação (3.27)

$$I_{\rho_\theta, v}(\Omega_i) = -\frac{1}{2} \text{Tr}([\Omega_i, \rho_\theta^v][\Omega_i, \rho_\theta^{1-v}]),$$

com $v \equiv (1 - \alpha)/2$ fazendo $v = [0, 1]$, onde $\rho = \log \rho$ se $v = 0, 1$.

Em particular, se calcularmos a norma $\langle \partial_i \ell^{(\alpha)}, \partial_i \ell^{(\alpha)} \rangle$ para $\alpha = 0$ obtemos (-8) vezes a informação de Wigner-Yanase⁶⁰ (WYSI)[§]

$$I_{\rho_\theta}(\Omega_i) = -\frac{1}{2} \text{Tr}([\Omega_i, \sqrt{\rho_\theta}]^2),$$

que é contrativa e apresenta α -conexão Riemanniana, pela extensão quântica do Teo. 3.

[†]Sigla vinda do inglês *completely positive trace preserving* que será utilizada para nos referirmos a estes mapas gerais.

[§]Vamos utilizar ao longo do texto a sigla WYSI, do inglês *Wigner-Yanase skew information*, para nos referirmos a informação de Wigner-Yanase.

4 O TEOREMA DE MCP E A RELAÇÃO ENTRE OPERADOR FUNÇÃO MONOTÔNICA E CONVEXA

Da seção Sec. 3.6.4, sabemos que para uma métrica ser considerada permitida em espaço de estados quânticos ela deve satisfazer a condição de contratividade sob da ação de mapas CPTP. A dedução da expressão destas métricas contrativas, Riemannianas e invariantes sob ação unitária, foi iniciada por Morozova e Chentsov²⁷ e finalizada por Petz²⁹ estabelecendo uma relação biunívoca entre métricas contrativas e funções monotônicas. Neste capítulo queremos apresentar estas métricas e observar, sob certas circunstâncias, como elas se relacionam com aquelas vistas no capítulo anterior para o caso quântico.

A Sec. 4.1 esta destinada a enunciarmos o Teorema de Morozova-Chentsov-Petz (MCP) que define as tais métricas quânticas permitidas. Seção esta baseada na tese de Pires.²³

Na Sec. 4.2, vamos calcular os coeficientes do tensor métrico quando o estado está na decomposição espectral dada uma base ortonormal para o espaço de Hilbert. Nos restringindo a dinâmica unitária, mostraremos um caso específico de função monotônica que resulta em coeficientes da métrica iguais aos vistos na (3.27) para $\alpha = 0$.

Por fim, na Sec. 4.3, seguindo o texto de Jenčova,⁶¹ apresentaremos uma identificação entre operadores funções monotônica e convexa que irá mostrar que, de fato, há uma relação entre as $g^{(\alpha)}$ métricas com as métricas vindas do Teorema MCP.

4.1 O Teorema de MCP

Seja o espaço de Hilbert $\mathcal{H}$, com $\dim \mathcal{H} = d$, e $\mathcal{S}$ o conjunto convexo de operadores densidade ρ hermitianos, positivos definidos de traço unitário em $\mathcal{H}$. Como sabemos, $\mathcal{S}$ é variedade diferenciável $(d^2 - 1)$ -dimensional. O espaço tangente a um ponto $\rho \in \mathcal{S}$, denotado por $T_\rho(\mathcal{S})$ é identificado com o espaço $\mathcal{A}_0$ de matrizes hermitianas de traço nulo. Vide Sec. 3.6.2

Para todo $X, Y \in T_\rho(\mathcal{S})$ e os operadores lineares $L_\rho, R_\rho : \mathcal{B} \rightarrow \mathcal{B}$ cuja ação é dada por

$$L_\rho A = \rho A, \quad R_\rho A = A \rho \quad \text{para todo } A \in \mathcal{B}, \quad (4.1)$$

onde $\mathcal{B}$ é o espaço dos operadores em $\mathcal{H}$, definimos em $\mathcal{S}$ uma métrica Riemanniana, $g_\rho^f : T_\rho(\mathcal{S}) \times T_\rho(\mathcal{S}) \rightarrow \mathbb{R}$ para todo $\rho \in \mathcal{S}$, dada por^{27,29}

$$g_\rho^f(X, Y) = \text{Tr} \left[X J_f^{-1}(L_\rho, R_\rho) Y \right] \quad (4.2)$$

com

$$J_f^{-1}(x, y) \equiv \frac{1}{y f(xy^{-1})},$$

sendo

- i) Positiva, $g_\rho^f(X, Y) \geq 0$.
- ii) Simétrica, $g_\rho^f(X, Y) = g_\rho^f(Y, X)$.
- iii) Invariante sobre ação de qualquer transformação unitária $U \in \mathcal{U}$,

$$g_\rho^f(UXU^\dagger, UYU^\dagger) = g_\rho^f(X, Y),$$

onde $\mathcal{U}$ é o grupo unitário contido em $\mathcal{B}$.

- iv) Suave em $\rho \in \mathcal{S}$ e $X, Y \in T_\rho(\mathcal{S})$.

- v) Contrativa sob ação de mapa CPTP $\epsilon(\cdot)$,

$$g_\rho^f(\epsilon(X), \epsilon(Y)) \leq g_\rho^f(X, Y).$$

As funções $f(u) : \mathbb{R}^+ \rightarrow \mathbb{R}$ pertencem a chamada classe de *funções de Morozova-Chentsov (MC)*, sendo

- i) Monotônicas, $u \leq v$ implica $f(u) \leq f(v)$.
- ii) Auto-inversa, $f(u) = uf(1/u)$.
- iii) Normalizada $f(1) = 1$.

Note que as propriedades i)+ii) das funções de MC implicam que $J_f^{-1}(x, y)$ é simétrica e satisfaz $J_f^{-1}(\alpha x, \alpha y) = \alpha^{-1} J_f^{-1}(x, y)$. Além disso estas funções respeitam uma hierarquia, de forma que

$$\frac{2u}{1+u} = f_{\min} \leq f(u) \leq f_{\max} = \frac{1+u}{2}.$$

Elas também não possuem estrutura totalmente arbitrária. Hansen^{62,63} mostrou que $f(u)$ tem representação canônica dada por

$$f(u) = \frac{(1+u)e^\vartheta}{\sqrt{2}} \exp \left[\int_0^1 \frac{(v^2-1)(1+u^2)j(v)}{(v^2+1)(v+u)(1+vu)} dv \right],$$

onde $j(v) : [0, 1] \rightarrow [0, 1]$ e $\vartheta \in \mathbb{R}$ são determinados univacamente pelas propriedades da função de MC. Em sua tese, Pires²³ faz uma discussão sobre como estas representações nos fornecem funções de MC de interesse.

Teorema 10. (Morozova, Chentsov, Petz) Considere $\rho \in \mathcal{S}$ e $X, Y \in T_\rho(\mathcal{S})$. Seja $g_\rho^f(\cdot, \cdot)$ tensor que garante a correspondência biunívoca entre métricas Riemannianas contrativas e a classe funções monotônicas de MC. Fixando uma base ortonormal $\{|k\rangle\}_{k=1,\dots,d}$ de $\mathcal{H}$

correspondente a decomposição espectral $\rho = \sum_{k=1}^d p_k |k\rangle \langle k|$, então a expressão da métrica no espaço de estados é dada por

$$g_\rho^f(X, Y) = \left[\sum_{i=1}^d \frac{X_{(ii)} Y_{(ii)}}{p_i} + 2 \sum_{\substack{i,j=1 \\ i < j}}^d J_f^{-1}(p_j, p_i) X_{(ij)} Y_{(ji)} \right],$$

onde $X_{(ij)} \equiv \langle i|X|j\rangle$ e $Y_{(ij)} \equiv \langle i|Y|j\rangle$.

Prova. Dado o estado diagonal não degenerado $\rho = \sum_{k=1}^d p_k |k\rangle \langle k|$ na base $\{|k\rangle\}_{k=1,\dots,d}$, vamos identificar $\mu_k := |k\rangle \langle k|$. Note que, $R_\rho^{-1} = R_{\rho^{-1}}$ e $L_\rho^{-1} = L_{\rho^{-1}}$, portanto o $L_\rho R_\rho^{-1}$ é igual ao operador relativo modular $\Delta_{\rho,\rho}$, dado na Eq.(3.23). Aplicando a propriedade (3.24) temos que

$$\Delta_{\rho,\rho}(\mu_j Y \mu_i) = \frac{p_j}{p_i} \mu_j Y \mu_i \quad \text{e} \quad f(\Delta_{\rho,\rho})(\mu_j Y \mu_i) = f\left(\frac{p_j}{p_i}\right) (\mu_j Y \mu_i).$$

Então $f(\Delta_{\rho,\rho})Y = \sum_{i,j} f(p_j/p_i)(\mu_j Y \mu_i)$, desta forma

$$R_\rho f(L_\rho R_\rho^{-1})Y = \sum_{i,j} p_i f\left(\frac{p_j}{p_i}\right) (\mu_j Y \mu_i).$$

Sabendo que R_ρ é operador hermitiano positivo, podemos aplicar novamente o raciocínio acima obtendo

$$[R_\rho f(L_\rho R_\rho^{-1})]^{-1}Y = J_f^{-1}(L_\rho, R_\rho)Y = \sum_{i,j} J_f^{-1}(p_j, p_i)(\mu_j Y \mu_i).$$

Calculando a expressão da métrica

$$\begin{aligned} g_\rho^f(X, Y) &= \text{Tr}[X J_f^{-1}(L_\rho, R_\rho)Y] \\ &= \sum_{ijk} J_f^{-1}(p_j, p_i) \langle k|X\mu_j Y \mu_i|k\rangle \\ &= \sum_{ij} J_f^{-1}(p_j, p_i) \langle i|X|j\rangle \langle j|Y|i\rangle. \end{aligned}$$

Como $J_f^{-1}(p_i, p_i) = 1/p_i$, separando a parte diagonal da parte não-diagonal, temos a expressão desejada

$$g_\rho^f(X, Y) = \left[\sum_{i=1}^d \frac{\langle i|X|i\rangle \langle i|Y|i\rangle}{p_i} + 2 \sum_{\substack{i,j=1 \\ i < j}}^d J_f^{-1}(p_j, p_i) \langle i|X|j\rangle \langle j|Y|i\rangle \right].$$

Uma outra prova deste fato pode ser vista na tese de Pires.²³

□

Um fato muito importante deve ser apontado. Já no capítulo anterior, trabalhamos apenas com o caso de $\text{rank}(\rho) = \dim \mathcal{H}$ com espectro não degenerado, evitando estados puros. Estas restrições podem se justificada olhando para a expressão do Teo. 10, pois

estados puros produzem em sua forma diagonal p_j 's nulos e espectro degenerado podendo fazer J_f^{-1} divergir, de forma que a métrica de MCP não vai ser bem definida. Já no capítulo anterior poderíamos ter apontado esse fato dada a porção da métrica (3.28). Portanto, a passagem do caso misto para o puro pode ser perigosa e não trivial.

Em sua tese, Pires comenta que MCP pode ser estendido ao caso puro sobre algumas circunstâncias, processo conhecido como extensão radial,⁶⁴ requerendo $f(0) \neq 0$, onde a pureza do estado faz com que as métricas colapsem na métrica de Fubini-Study, que é a métrica natural do espaço de estados puros, como já mencionamos no Cap. 3.

4.2 MCP, dinâmica unitária e Wigner-Yanase

4.2.1 Coeficientes da métrica

Seja um sistema quântico inicialmente descrito pelo estado inicial ρ_0 , sua evolução para um estado ρ_θ é descrita via transformação

$$\epsilon_\theta(\rho_0) = \rho_\theta,$$

onde $\epsilon_\theta(\cdot)$ é operação unitária ou não, rotulada por um conjunto de parâmetros $\theta = \{\theta^\mu\}_{\mu=1,\dots,r}$ e que, assim como ρ_θ , é suave em $\theta \in \mathbb{R}^r$. Temos que $[\theta^\mu]$ forma um sistema de coordenadas locais na variedade $\mathcal{S}$, e um elemento da base natural de $X \in T_\theta(\mathcal{S})$ será dado por $\partial_\mu \rho_\theta$, com $\partial_\mu = \partial/\partial\theta^\mu$.

Vamos então proceder com o cálculo dos coeficientes da métrica (4.2) nas condições do Teo. 10. Dado $\rho_\theta = \sum_{k=1}^d p_k |k\rangle \langle k|$ na base ortonormal $\{|k\rangle\}_{k=1,\dots,d}$ de $\mathcal{H}$, por hipótese temos $p_k = p_k(\theta)$ e $|k\rangle = |k(\theta)\rangle$. Um resultado auxiliar útil é que $\partial_\mu(\langle k|l\rangle) = \partial_\mu(\delta_{kl}) = 0 = (\partial_\mu \langle k|) |l\rangle + \langle k| (\partial_\mu |l\rangle)$, ou seja

$$(\partial_\mu \langle k|) |l\rangle = -\langle k| (\partial_\mu |l\rangle).$$

A expressão de um elemento da base de $T_\theta(\mathcal{S})$, em termos da base ortonormal de $\mathcal{H}$, é dada por

$$\partial_\mu \rho_\theta = \sum_{k=1}^d \left\{ (\partial_\mu p_k) |k\rangle \langle k| + p_k [(\partial_\mu |k\rangle) \langle k| + |k\rangle (\partial_\mu \langle k|)] \right\}.$$

Desta forma, o elemento de matriz $(\partial_\mu \rho_\theta)_{(ij)}$ será dado por

$$\langle i | \partial_\mu \rho_\theta | j \rangle = \delta_{ij} \partial_\mu p_i + (p_j - p_i) D_\mu^{(ij)},$$

onde definimos $D_\mu^{(ij)} \equiv \langle i | \partial_\mu | j \rangle$, o que nos leva a

$$\begin{aligned} (\partial_\mu \rho_\theta)_{(ii)} (\partial_\nu \rho_\theta)_{(ii)} &= \partial_\mu p_i \partial_\nu p_i, \\ (\partial_\mu \rho_\theta)_{(ij)} (\partial_\nu \rho_\theta)_{(ji)} &= -(p_i - p_j)^2 D_\mu^{(ij)} D_\nu^{(ji)}. \end{aligned}$$

Substituindo estes resultados no Teo. 10, vemos que a métrica contrativa associada a família de funções de MC se quebra como as métricas (3.30)

$$g_{\rho_\theta}^f(\partial_\mu \rho_\theta, \partial_\nu \rho_\theta) = g_{\mu,\nu}^f(\theta) = \mathcal{F}_{\mu\nu} + \mathcal{Q}_{\mu\nu}^f, \quad (4.3)$$

onde

$$\mathcal{F}_{\mu\nu} = \sum_{i=1}^d \frac{\partial_\mu p_i \partial_\nu p_i}{p_i}$$

é exatamente a expressão (3.29) correspondente a Fisher clássica, e

$$\mathcal{Q}_{\mu\nu}^f = -2 \sum_{\substack{i,j=1 \\ i < j}}^d J_f^{-1}(p_j, p_i) (p_i - p_j)^2 D_\mu^{(ij)} D_\nu^{(ji)}$$

corresponde as contribuições das coerências de $\partial_\mu \rho_\theta$, o análogo deste contexto a expressão (3.27). Da mesma forma que na expressão (3.30), ρ_θ e $\partial_\mu \rho_\theta$ comutarem implica que $g_{\mu,\nu}^f(\theta)$ se reduz a $\mathcal{F}_{\mu\nu}$.

4.2.2 Métrica de Wigner-Yanase na dinâmica unitária

Dizemos que a evolução do estado inicial ρ_0 do sistema é *unitária* quando é descrita via ação de um operador $U_\theta \in \mathcal{U}$, sendo $\mathcal{U}$ o grupo unitário, rotulado pelos parâmetros $\theta = \{\theta^\mu\}_{\mu=1,\dots,r}$ e dada por

$$\epsilon(\rho_0) = U_\theta \rho_0 U_\theta^\dagger.$$

Definimos os observáveis neste caso como

$$H_\mu(\theta) = iU_\theta^\dagger \partial_\mu U_\theta.$$

Seja a base $\{|k_0\rangle\}_{k=1,\dots,d}$ de $\mathcal{H}$ na qual o estado inicial tem a decomposição espectral

$$\rho_0 = \sum_{k=1}^d p_k |k_0\rangle \langle k_0|.$$

Agora vamos dizer que p_j e $|k_0\rangle$ não dependem do conjunto de parâmetros θ , ou seja, $\partial_\mu p_k = 0$ e $\partial_\mu |k_0\rangle = 0$. Definindo $|k\rangle \equiv U_\theta |k_0\rangle$, temos que

$$\rho_\theta = U_\theta \rho_0 U_\theta^\dagger = \sum_{k=1}^d p_k U_\theta |k_0\rangle \langle k_0| U_\theta^\dagger = \sum_{k=1}^d p_k |k\rangle \langle k|,$$

portanto, evolução unitária não altera a pureza e os estados inicial e final compartilham os mesmos autovalores. Consequentemente

$$\mathcal{F}_{\mu\nu} = 0.$$

Da relação $U_\theta U_\theta^\dagger = \mathbf{1}$, definimos

$$\Omega_\mu \equiv U_\theta^\dagger (\partial_\mu U_\theta) = -U_\theta (\partial_\mu U_\theta^\dagger) = -iH_\mu.$$

Sabendo que $U_\theta \Omega_\mu U_\theta^\dagger = \Omega_\mu$, podemos então calcular que

$$D_\mu^{(ij)} = \langle i | \partial_\mu | j \rangle = \langle i_0 | U_\theta^\dagger (\partial_\mu U_\theta) | j_0 \rangle = \langle i_0 | U_\theta \Omega_\mu U_\theta^\dagger | j_0 \rangle = \langle i | \Omega_\mu | j \rangle.$$

Aplicando estes resultados a expressão (4.3), obtemos

$$g_{\mu\nu}^f = \mathcal{Q}_{\mu\nu}^f = -2 \sum_{\substack{i,j=1 \\ i < j}}^d J_f^{-1}(p_j, p_i) (p_i - p_j)^2 \langle i | \Omega_\mu | j \rangle \langle j | \Omega_\nu | i \rangle. \quad (4.4)$$

Os coeficientes acima dependem da evolução unitária em questão, mostrando que a métrica está sujeita ao processo físico que evolui o estado. Além disso, $g_{\mu\nu}^f$ depende também da escolha de uma função de MC, diferentes funções vão dar diferentes expressões da métrica.

Em particular, vamos escolher a função

$$f_{WY}(u) = \frac{1}{4}(\sqrt{u} + 1)^2 \rightarrow J_{f_{WY}}^{-1}(x, y) = \frac{4}{(\sqrt{x} + \sqrt{y})^2}$$

para calcular expressão que resulta para os coeficientes da métrica. Substituindo então $J_{f_{WY}}^{-1}(x, y)$ em (4.4), sabendo que $\sum_{i,j} = 2 \sum_{i < j}$, obtemos

$$\begin{aligned} g_{\mu\nu}^{WY} &= -8 \sum_{\substack{i,j=1 \\ i < j}}^d \left(\frac{p_i - p_j}{\sqrt{p_i} + \sqrt{p_j}} \right)^2 \langle i | \Omega_\mu | j \rangle \langle j | \Omega_\nu | i \rangle \\ &= -8 \sum_{\substack{i,j=1 \\ i < j}}^d (\sqrt{p_i} - \sqrt{p_j})^2 \langle i | \Omega_\mu | j \rangle \langle j | \Omega_\nu | i \rangle \\ &= -16 \sum_{\substack{i,j=1 \\ i < j}}^d (p_i - \sqrt{p_i p_j}) \langle i | \Omega_\mu | j \rangle \langle j | \Omega_\nu | i \rangle \\ &= -8 \sum_{i,j=1}^d (p_i - \sqrt{p_i p_j}) \langle i | \Omega_\mu | j \rangle \langle j | \Omega_\nu | i \rangle \\ &= -8C(\Omega_\mu, \Omega_\nu) \\ &= 4\text{Tr}([\Omega_\mu, \sqrt{\rho_\theta}][\Omega_\nu, \sqrt{\rho_\theta}]), \end{aligned} \quad (4.5)$$

em que definimos

$$C(\Omega_\mu, \Omega_\nu) = -\frac{1}{2}\text{Tr}([\Omega_\mu, \sqrt{\rho_\theta}][\Omega_\nu, \sqrt{\rho_\theta}]).$$

Os coeficientes acima são exatamente os mesmos da expressão da métrica (3.27) para $\alpha = 0$. Com $\mu = \nu$ temos que

$$C(\Omega_\mu, \Omega_\nu) = -\frac{1}{2}\text{Tr}([\Omega_\mu, \sqrt{\rho_\theta}]^2) = I_{\rho_\theta}(\Omega_\mu)$$

e de novo vemos (-8) vezes a WYSI.

4.3 Relação operador função monotônica e convexa

As componentes (4.5) levantam um questionamento acerca de uma correspondência entre as métricas $g^{(\alpha)}$ (3.30) e as métricas g^f (4.2). Nesta seção vamos apresentar uma relação entre funções monotônicas de MC $f(u)$ com funções convexas $h(u)$ que estabelecem um elo entre o teorema MCP e as α -métricas.

O conjunto de *operadores função monotônicas* de MC é definido por funções $f(u) : \mathbb{R}^+ \rightarrow \mathbb{R}$, com $f(u) \leq f(v)$ se $u \leq v$, $f(u) = uf(u^{-1})$, e $f(1) = 1$. Seguindo os passos de Jenčová,⁶¹ apresentaremos uma discussão com um série de afirmações e resultados que tem sua prova muito além do escopo desta dissertação, portanto serão apenas enunciados para nos encaminharmos rapidamente ao objetivo. Vale notar que o texto de Hiai-Mosonyi também trata do assunto.⁶⁵

Definimos por $\mathfrak{H}$ o conjunto de *operadores função convexa* composto por funções $h(u) : \mathbb{R}^+ \rightarrow \mathbb{R}$, com $h(1) = 0$ e $h''(1) = 1$. Todo operador função convexa com $h(1) = 0$ pode ser escrito na forma

$$h(u) = a(u-1) + b(u-1)^2 + c \frac{(u-1)^2}{u} + \int_0^\infty (u-1)^2 \frac{1+s}{u+s} d\mu(s), \quad (4.6)$$

onde $a, b, c \in \mathbb{R}$, com $b, c \geq 0$ e μ é uma medida finita em $(0, \infty)$. A constante a não interfere em nenhuma das estruturas que irão seguir, portanto duas funções que diferem por ela são tratadas como iguais.

Seja $\mathfrak{M}$ conjunto de todas as medidas μ em $(0, \infty)$, onde $\int d\mu = 1/2$. A expressão (4.6) de h estabelece uma correspondência biunívoca entre $\mathfrak{H}$ e $\mathfrak{M}$ com $c = \mu(\{0\})$ e $c = \mu(\{\infty\})$. Defina a transposta de h como $\hat{h}(u) = uh(u^{-1})$. Se $h \in \mathfrak{H}$ então $\hat{h} \in \mathfrak{H}$, a relação $h \mapsto \hat{h}$ induz um mapa $\mathfrak{M} \rightarrow \mathfrak{M}$ dado por $\mu \mapsto \hat{\mu}$, de forma que $d\hat{\mu}(s) = d\mu(s^{-1})$. Assim, dizemos que h é simétrica se $h = \hat{h}$, que implica $\mathfrak{H}_{sim} \subset \mathfrak{H}$. Definimos então uma relação de equivalência $\sim$ em $\mathfrak{H}$, de maneira que $h_1 \sim h_2 \iff h_1 + \hat{h}_1 = h_2 + \hat{h}_2$. O quociente $\mathfrak{H}/\sim$ é isomorfo a $\mathfrak{H}_{sim}$.

Definimos, para $h \in \mathfrak{H}$, a h -entropia relativa $H_h : \mathcal{A}^+ \times \mathcal{A}^+ \rightarrow \mathbb{R}$ como

$$H_h(\rho, \sigma) = \text{Tr} \left[\sqrt{\rho} h \left(\frac{L_\sigma}{R_\rho} \right) \sqrt{\rho} \right] \quad \text{com} \quad \rho, \sigma \in \mathcal{A}^+,$$

onde $\mathcal{A}^+$ denota o conjunto de matrizes complexas hermitianas positivas definidas atuando em um espaço de Hilbert $\mathcal{H}$ d -dimensional.

Lesniewski e Ruskai⁶⁶ afirmam que dado $h \in \mathfrak{H}$, a, b, c e μ como em (4.6), vale

$$\begin{aligned} H_h(\rho, \sigma) &= a \text{Tr}(\sigma - \rho) \\ &+ \text{Tr}(\sigma - \rho) \left\{ b\rho^{-1} + c\sigma^{-1} + \int_0^\infty \frac{1+s}{L_\sigma + sR_\rho} d\mu(s) \right\} (\sigma - \rho) \\ &= a \text{Tr}(\sigma - \rho) + \text{Tr}(\sigma - \rho) R_\rho^{-1} K \left(\frac{L_\sigma}{R_\rho} \right) (\sigma - \rho), \end{aligned}$$

onde

$$K(u) = \int_0^\infty \frac{1+s}{u+s} d\mu(s) = \frac{h(u) - a(u-1)}{(u-1)^2}, \quad \text{com} \quad K(1) = 1/2.$$

Seja $X, Y \in T_\rho(\mathcal{S})$, então

$$g_\rho^f(X, Y) = -\frac{\partial^2}{\partial s \partial t} H_h(\rho + sX, \rho + tY)|_{s,t=0} = \text{Tr} \left[X R_\rho^{-1} K_{sim} \left(\frac{L_\rho}{R_\rho} \right) (Y) \right],$$

onde

$$K_{sim} = K(u) + u^{-1}K(u^{-1}) = \frac{h(u) + \hat{h}(u)}{(u-1)^2}.$$

Esta expressão define uma métrica monotônica, com a função monotônica valendo $f = 1/K$. Por outro lado, dada uma métrica monotônica, fazendo $a = 0$, podemos colocar em correspondência

$$h(u) = \frac{(u-1)^2}{2f(u)},$$

de forma que a condição $h''(1) = 1$ é equivalente a condição de normalização $f(1) = 1$.

Aplicando este resultado, vamos ver onde a $f_{WY}(u)$ é levada no conjunto de operadores funções convexas. Assim

$$h_{WY}(u) = \frac{(u-1)^2}{2f_{WY}(u)} = \frac{2(u-1)^2}{(\sqrt{u}+1)^2}.$$

Substituindo esta função na expressão (3.25) obtemos

$$D_{h_{WY}}(\rho \parallel \sigma) = 2\text{Tr}[(\sqrt{\rho} - \sqrt{\sigma})^2] = D^{(0)}(\rho \parallel \sigma),$$

o análogo quântico da distância de Hellinger (3.20). Como esperado, escolher a função $h_{WY}(u)$ é equivalente a escolher $h^{(\alpha=0)}(u)$ em (3.18), que corresponde a métrica $g_{ij}^{(0)u}$ em (3.27).

Um outro exemplo é escolhermos a função de MC associada a métrica de Bogoliubov ou produto interno de Kubo-Mori⁶⁷

$$f_{Bg}(u) = \frac{u-1}{\log u}.$$

Temos que

$$h_{Bg} = \frac{(u-1)^2}{2f_{Bg}(u)} = \frac{\log u(u-1)}{2}.$$

Substituindo na expressão (3.25) obtemos

$$D_{h_{Bg}}(\rho \parallel \sigma) = \frac{1}{2} [D^{(1)}(\sigma \parallel \rho) + D^{*(1)}(\sigma \parallel \rho)] = D^{(-1)}(\rho \parallel \sigma) = D^{(1)}(\sigma \parallel \rho).$$

Portanto escolher $h_{Bg}(u)$ é equivalente a escolher $h^{(\alpha=\pm 1)}(u)$, que irá corresponder a métrica $g_{ij}^{(\pm 1)u}$.

Um fato importante a se destacar. A relação entre os operadores função monotônica e convexa que obtivemos nesta seção não tem como hipótese inicial que $\alpha = 3 + 2h'''(1)$, com $\alpha \in \mathbb{R}$. Por isso não é sempre verdade que um operador função monotônica de MC $f(u)$ terá um correspondente no conjunto $h^{(\alpha)}(u)$. Porém é verdade que se $h^{(\alpha)}(u)$ for operador função convexa, que pela discussão da Sec. 3.6.4 é válido para $\alpha \in [-3, 3]$, ele terá um correspondente nos operadores função monotônica de MC.

5 CORREÇÕES NA COTA DE CRAMÉR–RAO

Neste capítulo, vamos dar uma formulação geométrica para a teoria de inferência estatística, clássica e quântica, apoiada na estrutura geométrica do espaço de Hilbert a fim de estabelecer ordens de correção superiores para a cota inferior da variância de um estimador de um único parâmetro. Ao longo do processo veremos aparecer intersecções dos nossos resultados com os formulados nos capítulos anteriores.

Especificamente, na Sec. 5.1 apresentaremos as operações estatísticas usuais numa roupaagem indexial que facilita explorarmos a geometria do espaço de funções densidade de probabilidade. Queremos estabelecer uma cota inferior para a variância de um estimador no problema de estimação de um parâmetro, observar como famílias exponenciais saturam a cota e estabelecer correções de ordem superior para a cota, chamadas cotas de Bhattacharyya.

Já na Sec. 5.2, vamos fazer a extensão para o caso quântico de estados puros, havendo a necessidade da introdução de uma estrutura complexa no espaço de Hilbert adjacente, notando que evolução do tipo Schrödinger não gera famílias de distribuição exponenciais no parâmetro de interesse, assim cotas inferiores do tipo Heisenberg não podem ser saturadas, portanto ordens superiores de correção serão importantes neste caso.

Por fim, na Sec. 5.3 queremos calcular as correções da cota no espaço das matrizes densidade. Para isso precisamos adaptar a discussão para uma notação análoga a das seções anteriores que comporte a álgebra de matrizes. Obteremos resultados que se assemelham na forma com os da Sec. 5.2, porém diferentes devido a não comutatividade do produto de matrizes.

A construção do raciocínio nas duas primeiras seções se dá seguindo os textos de Brody.^{3,68} Já na última seção usamos um outro texto do mesmo autor⁶⁹ para expor o problema e preparar o terreno para o cálculo das correções. O leitor interessado pode consultar estas referências para buscar informações não discutidas aqui.

5.1 Estatística clássica de um parâmetro

5.1.1 Estatística usual na notação indexial

Considere o espaço de Hilbert *real* $\mathcal{H}$, equipado com um produto interno *simétrico* g_{ab} . Repare que vamos trabalhar com a notação indexial para o espaço de Hilbert devido sua praticidade, onde, dada uma base ortonormal para $\mathcal{H}$, um típico vetor deste espaço pode ser representado como ξ^a . É importante destacar que no caso de dimensão infinita existem algumas tecnicidades matemáticas a serem levadas em conta, como o domínio dos operadores. É exatamente o problema de $\mathcal{X}$ infinito, discutido na Sec. C.2.

Especificando $\mathcal{H}$ como o espaço de todas as funções positivas $p(x)$ sobre o espaço amostral $\mathcal{X}$ na semi-reta $\mathbb{R}^+$. Tomando o mergulho $p \rightarrow \sqrt{p}$ (vide (3.9)) definimos o conjunto de todas as densidades de probabilidade $\mathcal{P}$ por

$$g_{ab}\xi^a\xi^b = \int_x \int_y \delta(x-y)p^{1/2}(x)p^{1/2}(y)dxdy = 1,$$

onde podemos fazer a identificação

$$\xi^a \rightarrow \sqrt{p(x)} ; g_{ab} \rightarrow \delta(x-y) ; a, b \rightarrow x, y. \quad (5.1)$$

Uma variável estocástica em $\mathcal{H}$ é representada por uma forma bilinear *simétrica* X_{ab} com valor esperado em termos do estado ξ^a dado por

$$E_\xi[X] = X_{ab}\xi^a\xi^b = \int_x \int_y x\delta(x-y)p^{1/2}(x)p^{1/2}(y)dxdy,$$

assim como em (5.1) identificamos $X_{ab} \rightarrow x\delta(x-y)$. Estendendo o procedimento para outras expressões bem conhecidas de estatística, como por exemplo o valor esperado do quadrado $X_{ab}X_{cd}\xi^a\xi^b\xi^c\xi^d$, definimos também o desvio da média $\tilde{X}_{ab} = X_{ab} - g_{ab}(X_{cd}\xi^c\xi^d)$, de onde obtemos

$$\text{Var}_\xi[X] = \tilde{X}_{ac}\tilde{X}_{bd}\xi^a\xi^b\xi^c\xi^d = X_{ac}X_{bd}\xi^a\xi^b\xi^c\xi^d - (X_{ab}\xi^a\xi^b)^2,$$

que nada mais é que a variância de X_{ab} no estado ξ^a . De forma análoga, a covariância entre as variáveis estocásticas X_{ab} e Y_{ab} no estado ξ^a pode ser reescrita como

$$\text{Cov}_\xi[X, Y] = \tilde{X}_{ac}\tilde{Y}_{bd}\xi^a\xi^b\xi^c\xi^d.$$

5.1.2 Métrica de Fisher

Considere agora que nossos vetores são parametrizados por um conjunto de *coordenadas locais* $\{\theta = [\theta^i]_{i=1,\dots,n} \in \Theta \subset \mathbb{R}^n\}$, desta forma temos uma subvariedade $\mathcal{S} \subseteq \mathcal{P}$ em $\mathcal{H}$ dada parametricamente por $\xi(\theta) \in \mathbb{C}^\infty$. Denotaremos ∂_i para $\partial/\partial\theta^i$.

Proposição 1. (*Métrica de Fisher*) Em coordenadas locais θ^i , a métrica Riemanniana $\mathcal{G}_{ij}$ em $\mathcal{S}$, induzida da métrica g_{ab} de $\mathcal{H}$, é dada pela matriz informação de Fisher

$$\mathcal{G}_{ij} = 4g_{ab}\partial_i\xi^a\partial_j\xi^b.$$

Prova. Note que a distância quadrada entre dois vetores é dada por $g_{ab}(\xi^a - \eta^a)(\xi^b - \eta^b)$. Se ambos os vetores pertencem a $\mathcal{S}$ e η^a é obtido por um deslocamento infinitesimal de $\xi^a \in \mathcal{S}$, ou seja, $\eta^a = \xi^a + \partial_i\xi^a d\theta^i$, então a separação ds entre os dois vetores em $\mathcal{S}$ é

$$d^2s = g_{ab}(\partial_i\xi^a d\theta^i)(\partial_j\xi^b d\theta^j) = g_{ab}\partial_i\xi^a\partial_j\xi^b d\theta^i d\theta^j \stackrel{(\times 4)}{=} \frac{1}{4}\mathcal{G}_{ij}d\theta^i d\theta^j.$$

□

O fator $1/4$ aparece em analogia a definição convencional da matriz informação de Fisher em termos da função $l(x, \theta) = \log p(x, \theta)$, dada por

$$\mathcal{G}_{ij} = \int p(x, \theta) \partial_i l(x, \theta) \partial_j l(x, \theta) dx = 4 \int \partial_i \sqrt{p(x, \theta)} \partial_j \sqrt{p(x, \theta)} dx.$$

Note que recuperamos as expressões (3.3) e (3.5) usando esse fator 4, que é exatamente o fator que aparece na métrica correspondente ao mergulho para $\alpha = 0$, $p \rightarrow 2\sqrt{p}$, em (3.10).

Diferenciando $g_{ab}\xi^a\xi^b = 1$ duas vezes, obtemos uma expressão alternativa para $\mathcal{G}_{ij}$

$$\begin{aligned} g_{ab}\xi^a\xi^b &= 1 \xrightarrow{\partial_i} g_{ab}[\partial_i\xi^a\xi^b + \xi^a\partial_i\xi^b] = 0 \rightarrow \\ g_{ab}\xi^a\partial_i\xi^b &= 0 \xrightarrow{\partial_j} g_{ab}[\partial_j\xi^a\partial_i\xi^b + \xi^a\partial_j\partial_i\xi^b] = 0 \xrightarrow{\times 4} \\ \mathcal{G}_{ij} &= 4g_{ab}\partial_j\xi^a\partial_i\xi^b = -4\xi^a\partial_j\partial_i\xi^b, \end{aligned} \quad (5.2)$$

análogo da Eq.(3.4).

A geometria induzida sobre $\mathcal{S}$ pode ser estudada em termos da métrica Riemanniana $\mathcal{G}_{ij}$, por isso vamos exemplificar alguns breves resultados nesse sentido.

Proposição 2. Os símbolos de Christoffel para a conexão Riemanniana vinda de $\mathcal{G}$ são dados por $\Gamma_{jk}^i = 4\mathcal{G}^{il}\partial_l\xi^a\partial_j\partial_k\xi_a$.

Prova. Uma variedade Riemanniana tem uma conexão natural que pode ser obtida inserindo a métrica de Fisher na relação (2.25)

$$\begin{aligned} \Gamma_{jk}^i &= 2\mathcal{G}^{il}g_{ab}(\partial_j\partial_k\xi^a\partial_l\xi^b + \partial_k\xi^a\partial_j\partial_l\xi^b + \partial_k\partial_j\xi^a\partial_l\xi^b + \partial_j\xi^a\partial_k\partial_l\xi^b \\ &\quad - \partial_l\partial_j\xi^a\partial_k\xi^b - \partial_j\xi^a\partial_l\partial_k\xi^b) \\ &= 4\mathcal{G}^{il}g_{ab}\partial_l\xi^b\partial_j\partial_k\xi^a = 4\mathcal{G}^{il}\partial_l\xi^b\partial_j\partial_k\xi_b. \end{aligned}$$

□

Definindo $(\omega^i)_a$ de forma que $(\omega^i)_a\partial_j\xi^a \equiv \delta_j^i \in \mathbb{R}$, podemos escrever as componentes da inversa da métrica de Fisher como

$$\mathcal{G}^{ij} = \frac{1}{4}g^{ab}(\omega^i)_a(\omega^j)_b,$$

onde agora $\mathcal{G}^{ik}\mathcal{G}_{kj} = \delta_j^i$. Desta forma podemos calcular que

$$\begin{aligned} \Gamma_{kij} &= \Gamma_{ij}^k\mathcal{G}_{km} = 4g_{ab}\partial_k\xi^b\partial_i\partial_j\xi^a, \\ \Gamma_{kij} &= \int dx \left(\frac{1}{p_\theta}\partial_k p_\theta\partial_i\partial_j p_\theta - \frac{1}{2p_\theta^2}\partial_i p_\theta\partial_j p_\theta\partial_k p_\theta \right) \end{aligned}$$

recuperando, dado o mergulho $p \rightarrow \sqrt{p}$, a expressão (3.7) que é a conexão Riemanniana para $\alpha = 0$.

Proposição 3. Se ∇_i denota o operador derivada covariante, com respeito a conexão Riemanniana da Prop. 2, ao longo do vetor tangente ∂_i , então um simples cálculo nos leva que

$$\mathcal{G}_{ij} = -4\xi_a \nabla_i \nabla_j \xi^a. \quad (5.3)$$

Prova. $\nabla_j \xi^a = \partial_j \xi^a \xrightarrow{\nabla_i} \partial_i \partial_j \xi^a - \Gamma_{ij}^k \partial_k \xi^a \xrightarrow{-4\xi_a} -4\xi_a \partial_i \partial_j \xi^a$. Agora basta aplicarmos (5.2) para obtermos o resultado pretendido. $\square$

Poderíamos nos perguntar se existem outras conexões Riemannianas associadas a estrutura de $\mathcal{H}$. Para isso precisaríamos construir tensor do tipo (0,3) apenas utilizando a métrica e derivadas covariantes de ξ^a com respeito a conexão Riemanniana. Esta questão é relevante desde que quiséssemos saber se é possível construir um conjunto de conexões afins – ao exemplo de conexões- α em Amari (3.6) – puramente em termos da geometria base do espaço de Hilbert. Claramente, os únicos candidatos possíveis são $\nabla_i \xi^a \nabla_j \nabla_k \xi_a$ e $\xi^a \nabla_i \nabla_j \nabla_k \xi_a$. Entretanto, o seguinte resultado nos dá algo interessante sobre as conexões.

Proposição 4. As expressões $\nabla_i \xi^a \nabla_j \nabla_k \xi_a$ e $\xi^a \nabla_i \nabla_j \nabla_k \xi_a$ se anulam. Desta forma, nenhum outro tensor (0,3) pode ser construído de forma natural em $\mathcal{H}$, e a conexão Riemanniana distingue-se entre as possíveis α -conexões.

Prova. Primeiramente, sabemos que conexão Riemanniana é livre de torção e tem $\nabla_k \mathcal{G}_{ij} = 0$ por definição, implicando, via (5.2), que $\nabla_k (\nabla_i \xi^a \nabla_j \xi_a) = \nabla_k \nabla_i \xi^a \nabla_j \xi_a + \nabla_k \nabla_j \xi^a \nabla_i \xi_a = 0$. Diferenciando (5.3) obtemos $\nabla_k \xi_a \nabla_i \nabla_j \xi^a = -\xi_a \nabla_k \nabla_i \nabla_j \xi^a$. Desta forma, temos que $\nabla_k \nabla_i \xi^a \nabla_j \xi_a = \xi_a \nabla_i \nabla_j \nabla_k \xi^a$, e que $\xi_a \nabla_{(i} \nabla_{j)} \nabla_k \xi^a = 0$, onde os parêntesis indicam simetrização nos índices i e j . O termo $\xi_a \nabla_i \nabla_j \nabla_k$ é antissimétrico sobre os índices i, j . Definimos então o tensor de Riemann para qualquer campo vetorial suave covariante V_k como $(\nabla_i \nabla_j - \nabla_j \nabla_i) V_k = R_{ijk}{}^l V_l$. Segue que

$$\xi_a \nabla_i \nabla_j \nabla_k = \frac{1}{2} \xi_a R_{ijk}{}^l \nabla_l \xi^a.$$

Entretanto, $\xi_a \nabla_l \xi^a = 0$, visto que $\xi^a \xi_a = 1$, estabelecendo o resultado desejado. $\square$

Concluimos que *do ponto de vista do espaço de Hilbert, a conexão vista na Prop. 2 é a única conexão natural*. Esse fato já é esperado, afinal da Sec. 2.4.4 sabemos que a conexão Riemanniana definida por um tensor métrico é única.

5.1.3 Cota de Cramér–Rao e famílias exponenciais

Suponha que agora nós temos o resultado de uma medida do observável X_{ab} e que uma *família uniparamétrica* $\xi^a(\theta)$ de estados caracteriza a distribuição dos possíveis resultados da medida. Famílias uniparamétricas de distribuições fazem a métrica de Fisher se reduzir-se a

$$\mathcal{G} = 4\dot{\xi}^a \dot{\xi}_a = 4g_{ab} \dot{\xi}^a \dot{\xi}^b = 4 \left\| \dot{\xi}^a \right\|^2,$$

onde o ponto denota $\partial/\partial\theta$. Desejamos estimar θ com o resultado dado, já que θ determina o estado desconhecido. Para tal, uma cota inferior pode ser estabelecida para a variância com que a estimativa se desvia do valor real do parâmetro. Comparando a variância da nossa estimativa com a cota inferior, podemos nos perguntar o quanto nossa estimativa é boa. Vamos então derivar esta cota.

Teorema 11. (Desigualdade de Cramér-Rao) Seja T_{ab} um estimador não viesado *simétrico* para uma função $\tau(\theta)$, de forma que $T_{ab}\xi^a(\theta)\xi^b(\theta) = \tau(\theta)$, onde θ parametriza uma família uniparamétrica de estados $\xi^a(\theta)$ em $\mathcal{P} \subset \mathcal{H}$. Então a cota inferior da variância no estado ξ^a é dada por

$$\text{Var}_\xi[T] \geq \frac{\dot{\tau}^2}{4\dot{\xi}^a\dot{\xi}_a}. \quad (5.4)$$

Prova. Defina o desvio da média do estimador como $\tilde{T}_{ab} \equiv T_{ab} - \tau g_{ab}$, então $\tilde{T}_{ab}\xi^a\xi^b = 0$ e, $\text{Var}_\xi[T] = \tilde{T}_{ac}\tilde{T}_b^c\xi^a\xi^b$. Agora, defina $\eta_b \equiv \tilde{T}_{ab}\xi^a$ e note que

$$g_{ab}\eta^a\eta^b = \tilde{T}_c^a\xi^c g_{ab}\tilde{T}_d^b\xi^d = \tilde{T}_{da}\tilde{T}_c^a\xi^d\xi^c = \text{Var}_\xi[T].$$

Dado que $T_{ab}\xi^a\xi^b = \tau$ e a condição de normalização de ξ^a diante de g_{ab} , obtemos que

$$\begin{aligned} T_{ab}\xi^a\xi^b &= \tau \xrightarrow{\partial/\partial\theta} T_{ab}\xi^a\dot{\xi}^b = \dot{\tau}/2, \\ g_{ab}\xi^a\xi^b &= 1 \xrightarrow{\partial/\partial\theta} g_{ab}\xi^a\dot{\xi}^b = 0, \end{aligned} \quad (5.5)$$

e portanto

$$\eta_b\dot{\xi}^b = (T_{ab} - \tau g_{ab})\xi^a\dot{\xi}^b = \dot{\tau}/2.$$

Evocando a desigualdade de Cauchy-Schwartz $(\eta^a\eta_a)(\dot{\xi}^a\dot{\xi}_a) \geq (\eta_b\dot{\xi}^b)^2$, obtendo assim a expressão da desigualdade de Cramér-Rao. $\square$

O lado direito da desigualdade é denotada, *cota de Cramér-Rao (CRB) para a variância do estimador*, note também que o denominador do lado direito da expressão nada mais é do que a informação de Fisher $\mathcal{G}$, ou seja, quanto maior a informação que o estado prove sobre o parâmetro melhor será a estimativa.

É claro também que a cota é saturada na medida que a desigualdade de Cauchy-Schwartz é saturada, isto é se $\dot{\xi}^a = c\eta^a$ para alguma constante c que, dada uma reescala do parâmetro θ , pode ser fixada como $c = 1/2$ sem perda de generalidade. Desta forma, uma curva $\xi^a(\theta)$ que satura a cota deve satisfazer a equação diferencial

$$\dot{\xi}^a = \frac{1}{2}\tilde{T}_b^a\xi^b, \quad (5.6)$$

cuja solução é dada por uma *família exponencial canônica*⁷⁰ de distribuições com a seguinte expressão

$$\xi^a(\theta) = \frac{\exp\left[\frac{1}{2}\theta T_b^a\right] q^b}{\sqrt{\exp[\theta] T_b^a q^b q_a}}, \quad (5.7)$$

onde o vetor normalizado $\xi^a(0) = q^a/(q_b q^b)^{1/2}$ descreve a distribuição inicial, também sem perda de generalidade podemos escolher $q_a q^a = 1$. (vide apêndice C)

A expressão (5.7) nos leva a uma interessante interpretação geométrica das famílias exponenciais. Considere $\mathcal{P} \subset \mathcal{H}$ com a geometria induzida por g_{ab} , seja a forma

$$\tau(\xi^a) = \frac{T_{ab} \xi^a \xi^b}{g_{cd} \xi^c \xi^d}$$

definida em $\mathcal{H}$. Podemos então folhear $\mathcal{P}$ por superfícies a τ constante, Fig. 4. O gradiente da função $\tau(\xi^a)$, onde $\nabla_c = \partial/\partial \xi^c = \partial_c$ e $\partial_c \xi^a = \delta_c^a$, é dado por

$$\begin{aligned} \nabla_c \tau &= \frac{T_{ab}[(\partial_c \xi^a) \xi^b] + \xi^a (\partial_c \xi^b)}{\xi^d \xi_d} - \frac{T_{ef} \xi^e \xi^f}{(\xi^d \xi_d)^2} g_{ab}[(\partial_c \xi^a) \xi^b] + \xi^a (\partial_c \xi^b)] \\ &= \frac{2(T_{cb} - \tau g_{cb}) \xi^b}{\xi^d \xi_d} = \frac{2\tilde{T}_{cb} \xi^b}{\xi^d \xi_d} \\ \nabla^c \tau &= g^{ca} \nabla_a \tau = \frac{2\tilde{T}_b^c \xi^b}{\xi^d \xi_d}. \end{aligned} \quad (5.8)$$

Proposição 5. A família exponencial canônica de distribuições (5.7), com a distribuição inicial q^a , é dada pela única curva que passa pelo ponto q^a que é ortogonal em todo ponto a família de superfícies τ que folheiam $\mathcal{P}$.

Figura 4 – $\mathcal{P} \in \mathcal{H}$ folheado por superfícies a τ constante. Para uma família uniparamétrica qualquer $\dot{\xi}^a$ e $\nabla^a \tau$ não são paralelos; apenas famílias exponenciais, geradas a partir de um ponto q , cruzam perpendicularmente as superfícies τ .

Prova. Desde que, de acordo com (5.8), o vetor tangente (5.6) é paralelo ao gradiente da função τ em $\mathcal{P}$, temos o resultado proposto. $\square$

Proposição 6. Seja $\nabla_a = \partial/\partial\xi^a$ o operador gradiente em $\mathcal{H}$. Então, a variância do estimador não viesado T_{ab} da função τ no estado ξ^a é dada, em $\mathcal{P}$, pela expressão

$$\text{Var}_\xi[T] = \frac{1}{4} \nabla_a \tau \nabla^a \tau.$$

Prova. De acordo com as expressões em (5.8), segue que

$$\begin{aligned} \nabla_a \tau \nabla^a \tau &= \|\nabla^a \tau\|^2 = \left[\frac{2(T_{ab}\xi^b - \tau\xi_a)}{\xi^c \xi_c} \right] \left[\frac{2(T^a_d \xi^d - \tau\xi^a)}{\xi^c \xi_c} \right] \\ &= \frac{4}{(\xi^c \xi_c)^2} (T_{ba} T^a_d \xi^b \xi^d - \tau \xi_a T^a_d \xi^d - \tau T_{ab} \xi^a \xi^b + \tau^2 \xi^a \xi_a) \\ &= \frac{4}{(\xi^c \xi_c)^2} (T_{ba} T^a_d \xi^b \xi^d + [-\tau^2 - \tau^2 + \tau^2] \xi^a \xi_a) \\ &= \frac{4(T_{ba} T^a_d \xi^b \xi^d - \tau^2 g_{bd}) \xi^b \xi^d}{(\xi^c \xi_c)^2}, \end{aligned}$$

dado que a variância do estimador T_{ab} pode ser escrita como $\text{Var}_\xi[T] = (T_{ac} T^c_b - \tau^2 g_{ab}) \xi^a \xi^b / \xi^d \xi_d$, a expressão pretendida é obtida dividindo a norma do gradiente por 4 e restringindo ξ^a à $\mathcal{P}$, ou seja, $\xi^a \xi_a = 1$. $\square$

Em vista dos resultados (5.5) e (5.8), podemos calcular o produto interno restrito a $\mathcal{P}$ dos vetores $\dot{\xi}^a$ e $\nabla^a \tau$

$$\begin{aligned} g_{ab} \nabla^a \tau \dot{\xi}^b &\stackrel{\xi^a \in \mathcal{P}}{=} 2g_{ab} (T^a_c \xi^c - \tau \xi^a) \dot{\xi}^b \\ &= 2(T_{bc} \xi^c \dot{\xi}^b - \tau g_{ab} \xi^a \dot{\xi}^b) = \dot{\tau}, \end{aligned}$$

utilizando novamente a desigualdade de Cauchy-Schwartz recuperamos a expressão do CRB utilizando apenas entes geométricos definidos em $\mathcal{P}$,

$$\begin{aligned} |g_{ab} \nabla^a \tau \dot{\xi}^b|^2 &\leq \|\nabla^a \tau\|^2 \|\dot{\xi}^a\|^2 \\ \dot{\tau}^2 &\leq 4 \text{Var}_\xi[T] \dot{\xi}^a \dot{\xi}_a \rightarrow \text{Var}_\xi[T] \geq \dot{\tau}^2 / \mathcal{G}. \end{aligned}$$

Resumindo os resultados geométricos obtidos

- i) Pela definição de gradiente, $\nabla^a \tau$ é normal as hipersuperfícies a τ constante definidas em $\mathcal{P}$ que passam por ξ^a .
- ii) $\text{Var}_\xi[T] = (1/4) \|\nabla^a \tau\|^2$ no ponto ξ^a .
- iii) $\mathcal{G} = 4 \|\dot{\xi}^a\|^2$ no ponto ξ^a .
- iv) $g_{ab} \nabla^a \tau \dot{\xi}^b = \dot{\tau}$ implica no CRB, $\text{Var}_\xi[T] \geq \dot{\tau}^2 / \mathcal{G}$.

5.1.4 Cotas de Bhattacharyya

Acabamos de ver na seção anterior que as famílias exponenciais de distribuição (5.7) são as únicas capazes de saturar o CRB. Porém, para outros tipos de distribuições a variância irá exceder a cota. Neste caso, vamos considerar a possibilidade de estabelecer correções de ordem superior para as cotas. Nossa abordagem será relacionada aos trabalhos de Bhattacharyya.⁷¹⁻⁷³ Formularemos uma nova derivação da desigualdade de Cramér-Rao levando em conta aspectos geométricos de $\mathcal{P}$.

Para ilustrar a ideia, dado T_{ab} estimador não viesado para a função $\tau(\theta)$ então $\mathfrak{R}_{ab} = T_{ab} + \lambda \xi_{(a} \dot{\xi}_{b)}$ também o é para uma constante arbitrária λ . Naturalmente, a variância de $\mathfrak{R}_{ab}$ pode ser escrita como $\text{Var}_{\xi}[\mathfrak{R}] = \mathfrak{R}_{ac} \mathfrak{R}_b^c \xi^a \xi^b - (\mathfrak{R}_{ab} \xi^a \xi^b)^2$.

Tendo em mente o resultado de (5.5), vamos proceder com o cálculo da variância de $\mathfrak{R}_{ab}$ no estado ξ^a em $\mathcal{P}$, onde

$$\begin{aligned} \mathfrak{R}_{ab} \xi^a \xi^b &= (T_{ab} + \lambda [\xi_a \dot{\xi}_b + \xi_b \dot{\xi}_a]) \xi^a \xi^b = \tau \\ \mathfrak{R}_{ac} \mathfrak{R}_b^c \xi^a \xi^b &= \{(T_{ab} + \lambda [\xi_a \dot{\xi}_c + \xi_c \dot{\xi}_a])(T_b^c + \lambda [\xi^c \dot{\xi}_b + \xi_b \dot{\xi}^c])\} \xi^a \xi^b \\ &= T_{ac} T_b^c \xi^a \xi^b + \lambda \dot{\tau} + \dot{\tau}^2 \dot{\xi}^c \dot{\xi}_c \\ \text{Var}_{\xi}[\mathfrak{R}] &= \text{Var}_{\xi}[T] + \lambda \dot{\tau} + \dot{\tau}^2 \dot{\xi}^c \dot{\xi}_c. \end{aligned}$$

Queremos escolher o valor de λ que minimiza a variância acima. O procedimento é simples, já que a expressão é quadrática convexa em λ , então basta considerar a variância como função de λ e minimizá-la derivando na variável e igualando a zero, obtendo o valor $\lambda^{\min} = -\dot{\tau}^2 / 2 \dot{\xi}^a \dot{\xi}_a$. Desta maneira

$$\min_{\lambda}(\text{Var}_{\xi}[\mathfrak{R}]) = \text{Var}_{\xi}[T] - \frac{\dot{\tau}^2}{4 \dot{\xi}^a \dot{\xi}_a}.$$

Desde que $\text{Var}_{\xi}[\mathfrak{R}] \geq 0$, recuperamos imediatamente a expressão da desigualdade de Cramér-Rao.

Agora vamos estender a ideia acima para derivadas de ordem superior de $\xi^a(\theta)$. Denotamos então a r -ésima derivada de ξ^a com respeito ao parâmetro θ por $\xi^{(r)a} = d^r \xi^a / d\theta^r$ e definimos $\hat{\xi}^{(r)a}$ como a projeção de $\xi^{(r)a}$ ortogonal a ξ^a e à todas as derivadas de ordem menores que r

$$\hat{\xi}^{(r)a} = \xi^{(r)a} - \frac{\xi^{(r-1)b} \xi_b^{(r)}}{\xi^{(r-1)c} \xi_c^{(r-1)}} \xi^{(r-1)a} - \dots - (\xi^b \xi_b^{(r)}) \xi^a, \quad (5.9)$$

desta maneira

$$\begin{aligned} \hat{\xi}_a^{(r)} \xi^a &= 0 \\ \hat{\xi}_a^{(r)} \xi^{(s)a} &= 0; \quad r \neq s \\ \hat{\xi}_a^{(r)} \hat{\xi}^{(s)a} &= 0; \quad r \neq s, \end{aligned} \quad (5.10)$$

vamos considerar apenas valores de r onde $\xi^{(r)a} \neq 0$.

Se T_{ab} é um estimador não viesado da função $\tau(\theta)$ então o tensor simétrico

$$R_{ab} \equiv T_{ab} + \sum_r \lambda_r \xi_{(a} \hat{\xi}_{b)}^{(r)}$$

também o é, para um conjunto de constantes arbitrárias $\{\lambda_r\}_{r \in \mathbb{N}^\times}$. Seguindo com o cálculo da variância de R_{ab} no estado ξ^a em $\mathcal{P}$ obtemos

$$\begin{aligned} R_{ab} \xi^a \xi^b &= (T_{ab} + \sum_r \lambda_r \xi_{(a} \hat{\xi}_{b)}^{(r)}) \xi^a \xi^b = \tau \\ R_{ac} R_{b}^c \xi^a \xi^b &= \{(T_{ac} + \sum_r \lambda_r [\xi_a \hat{\xi}_c^{(r)} + \xi_c \hat{\xi}_a^{(r)}]) (T_b^c + \sum_s \lambda_s [\xi^c \hat{\xi}_b^{(s)} + \xi_b \hat{\xi}^{(s)c}])\} \xi^a \xi^b \\ &= T_{ac} T_b^c \xi^a \xi^b + \sum_s \lambda_s T_{ac} \xi^a \hat{\xi}^{(s)c} + \sum_r \lambda_r T_b^c \xi_b \hat{\xi}_c^{(r)} + \sum_{rs} \lambda_r \lambda_s \hat{\xi}_c^{(r)} \hat{\xi}^{(s)c} \\ \text{Var}_\xi[R] &= \text{Var}_\xi[T] + 2 \sum_r \lambda_r T_{ac} \xi^a \hat{\xi}^{(r)c} + \sum_r \lambda_r^2 \hat{\xi}_c^{(r)} \hat{\xi}^{(r)c}. \end{aligned}$$

Agora devemos determinar qual o valor que cada λ_r assume para minimizar a variância acima. Basta considerarmos a variância como função de variáveis $\{\lambda_r\}_{r \in \mathbb{N}^\times}$ e encontrar pontos extremos que anulam o gradiente desta função. Como a dependência da variância em cada um dos λ_r 's é a mesma, tendo formato quadrático convexa, sabemos que o ponto de extremo que produz um mínimo possui todas as entradas com o mesmo valor a saber $\lambda_r^{\min} = -T_{ac} \xi^a \hat{\xi}^{(r)c} / \hat{\xi}_c^{(r)} \hat{\xi}^{(r)c}$, para todo r . Desta forma obtemos

$$\min_{\{\lambda_r\}_{r \in \mathbb{N}^\times}} (\text{Var}_\xi[R]) = \text{Var}_\xi[T] - \sum_r \frac{(T_{ab} \xi^a \hat{\xi}^{(r)b})^2}{g_{ab} \hat{\xi}^{(r)a} \hat{\xi}^{(r)b}}.$$

Novamente, desde que $\text{Var}_\xi[R] \geq 0$, deduzimos as *cotas generalizadas de Bhattacharyya* para a variância do estimador T_{ab}

$$\text{Var}_\xi[T] \geq \sum_r \frac{(T_{ab} \xi^a \hat{\xi}^{(r)b})^2}{g_{ab} \hat{\xi}^{(r)a} \hat{\xi}^{(r)b}}. \quad (5.11)$$

Entretanto, dado o resultado da Prop. 6, podemos reescrever a desigualdade acima na em termos de entes geométricos de $\mathcal{P}$.

Proposição 7. As cotas inferiores generalizadas para a variância de um estimador não viesado T_{ab} de uma função τ pode ser expressa na forma

$$\text{Var}_\xi[T] \geq \frac{1}{4} \sum_r \frac{(\hat{\xi}^{(r)a} \nabla_a \tau)^2}{\hat{\xi}^{(r)b} \hat{\xi}_b^{(r)}}.$$

Prova. Para ξ^a restrito a $\mathcal{P}$ temos

$$\begin{aligned} \frac{1}{2} \nabla_a \tau &= T_{ab} \xi^b - (T_{cd} \xi^c \xi^d) \xi_a \\ \frac{1}{2} \hat{\xi}^{(r)a} \nabla_a \tau &= T_{ab} \hat{\xi}^{(r)a} \xi^b, \end{aligned}$$

inserindo esta última relação na expressão (5.11) chegamos no resultado de pretendido. Claramente, para $r = 1$ recobramos a desigualdade de Cramér–Rao. $\square$

A Prop. 7 nos diz que, dado o vetor gradiente $\nabla^a \tau \in \mathcal{H}$, o módulo quadrado deste vetor nada mais é do que a soma dos quadrados de suas componentes ortogonais com respeito a uma dada base de $\mathcal{H}$. No caso, a base é formada considerando ξ^a e suas derivadas de ordem superior que formam um conjunto ortonormal de vetores com respeito a g_{ab} dado por $\hat{\xi}^{(r)a} / (\hat{\xi}^{(r)b} \hat{\xi}_b^{(r)})^{1/2}$. Note que aplicarmos Cauchy-Schwartz anteriormente é o mesmo que usar Bhattacharyya para ordem 1, já evidenciando este fato. Essas novas cotas generalizadas *não* são necessariamente independentes do estimador T_{ab}

Por questão de exemplificação, vamos considerar o caso $r = 1, 2$, que segundo a Prop. 7 fornece

$$\text{Var}_\xi[T] \geq \frac{(\dot{\xi}^a \nabla_a \tau)^2}{4 \dot{\xi}^a \dot{\xi}_a} + \frac{(\hat{\xi}^{(2)a} \nabla_a \tau)^2}{\hat{\xi}^{(2)a} \hat{\xi}_a^{(2)}},$$

onde $\hat{\xi}^{(2)a} = \ddot{\xi}^a - \frac{(\ddot{\xi}^b \dot{\xi}_b)}{\dot{\xi}^c \dot{\xi}_c} \dot{\xi}^a - (\ddot{\xi}^b \xi_b) \xi^a$ é o vetor “aceleração”. Note que $\hat{\xi}^{(2)a} \hat{\xi}_a^{(2)} = (\dot{\xi}^a \dot{\xi}_a)^2 K_\xi^2$ onde K_ξ^2 é a curvatura intrínseca –“aquela vista por quem vive na variedade”– da curva $\xi^a(\theta)$.⁷⁴ Uma desigualdade estatística garante que $K_\xi^2 \geq 0$.⁷⁵ Desde que o segundo termo da expressão acima é sempre não negativo, enquanto o primeiro termo é o CRB, temos claramente uma correção na cota. Ganhamos então nossa interpretação geométrica: a desigualdade de Cramér–Rao melhorada acima nos diz que o comprimento do vetor $\nabla^a \tau$ é maior que sua projeção ortogonal no plano definido pelos vetores $\dot{\xi}^a$ e $\hat{\xi}^{(2)a}$. (vide apêndice C)

5.2 Estatística quântica de um parâmetro, caso puro

5.2.1 Geometria de estado puro

Agora avançamos para caso quântico. Primeiro temos que formular a teoria quântica padrão de estado puro em termos de uma linguagem geométrica para podermos estabelecer um paralelo com a geometria estatística desenvolvida na Sec. 5.1. No caminho até aqui, formulamos a estimativa clássica de um parâmetro baseados na geometria do espaço de Hilbert real. A fim de futuramente desenvolvermos dinâmica do tipo Schrödinger, devemos inserir uma *estrutura complexa compatível* ao espaço de Hilbert $\mathcal{H}$ real visto anteriormente, algo bem conhecido em textos de geometria complexa.^{50,76}

Uma estrutura complexa em $\mathcal{H}$ é definida por um tensor linear J_a^b satisfazendo

$$J_a^b J_b^c = -\delta_a^c.$$

Quando requeremos que a estrutura complexa tenha compatibilidade com a estrutura de $\mathcal{H}$ estamos pedindo que o produto interno g_{ab} seja *hermitiano* via relação

$$J_a^c J_d^b g_{ab} = g_{cd},$$

para todo vetor pertencente ao espaço de Hilbert. Definida esta estrutura, temos que qualquer operador hermitiano simétrico X_{ab} satisfaz a relação $J^a_c J^b_d X_{ab} = X_{cd}$, que é equivalente à $J^a_c X^c_b = X^a_c J^c_b$; isto segue da definição da estrutura complexa, da hermiticidade do produto interno e da anti-simetria do tensor $g_{ac} J^c_b$, que vamos tratar logo a frente.

Vamos trazer agora a abordagem para um lugar conhecido ao olhar físico, comparando a notação indexial com a notação de Dirac convencional. Em uma abordagem real da teoria quântica, vetores de estado são representados por elementos de $\mathcal{H}$. Temos então que se ξ^a e η^a são dois vetores reais de $\mathcal{H}$, então seu *produto de Dirac* será dado pela seguinte expressão

$$\langle \eta | \xi \rangle = \frac{1}{2} \eta^a (g_{ab} - i g_{ac} J^c_b) \xi^b. \quad (5.12)$$

A hermiticidade de g_{ab} implica que o tensor $\Omega_{ab} \equiv g_{ac} J^c_b$ seja automaticamente antissimétrico e invertível, ou seja, uma *estrutura simplética*, que também satisfaz a condição de hermiticidade $J^a_c J^b_d \Omega_{ab} = \Omega_{cd}$. Visto que a estrutura simplética Ω_{ab} é antissimétrica segue diretamente que a norma de Dirac é compatível com a norma real de Hilbert, a menos de um fator dois

$$\langle \xi | \xi \rangle = \frac{1}{2} g_{ab} \xi^a \xi^b.$$

Vamos definir o par de vetores

$$\xi_{\pm}^a \equiv \frac{1}{2} (\xi^a \mp i J^a_b \xi^b).$$

Note que

$$\begin{aligned} J^a_b \xi_{\pm}^b &= \frac{1}{2} (J^a_b \xi^b \mp i (-\delta^a_c) \xi^c) = \frac{1}{2} (J^a_b \xi^b \pm i \xi^a) \\ &= (\pm i) \frac{1}{2} (\mp J^a_b \xi^b + \xi^a) = \pm i \xi_{\pm}^a. \end{aligned}$$

Ou seja, ξ_+^a e ξ_-^a são autovetores complexos de J^a_b . O resultado da expressão acima nos permite decompor o espaço de Hilbert real de vetores ξ^a em parte complexa positiva e negativa relativas a estrutura complexa especificada, onde denotaremos $\xi^a = \xi_+^a + \xi_-^a$, dada a divisão do espaço de Hilbert $\mathcal{H}$ em autoespaço positivo e negativo como $\mathcal{H} = \mathcal{H}^+ \oplus \mathcal{H}^-$. Como consequência da condição de hermiticidade e antissimetria de Ω_{ab} , dois vetores do mesmo tipo são necessariamente ortogonais com respeito à métrica g_{ab} , em outros termos, temos que $g_{ab} \xi_{\pm}^a \eta_{\pm}^b = 0$ para qualquer dois vetores positivos/negativos $\xi_{\pm}^a$ e $\eta_{\pm}^a$.

Vamos utilizar então *índices gregos* para denotar a parte positiva e negativa a fim de justificar a escolha do produto de Dirac com a forma (5.12). Escrevendo $\xi^a = (\xi^\alpha, \bar{\xi}_\alpha)$, onde $\bar{\xi}_\alpha$ é o complexo conjugado de ξ^α . Os índices α em cima e embaixo querem denotar o fato de $\mathcal{H}^-$ ser o dual associado a $\mathcal{H}^+$, este sendo tratado como o espaço direto no caso. Podemos identificar ξ^α com o vetor $|\xi\rangle$ de Dirac, e o $\bar{\xi}_\alpha$ com o covetor $\langle \bar{\xi}|$ complexo conjugado de Dirac, escrevendo $\xi^a = (|\xi\rangle, \langle \bar{\xi}|)$. Assim, um vetor do espaço de Hilbert

complexo denotado como ψ^α , ou $|\psi\rangle$ na notação convencional de Dirac, e um elemento do dual denotado por $\phi_\alpha = \langle\phi|$ tem seu produto interno escrito como $\phi_\alpha\psi^\alpha = \langle\phi|\psi\rangle$. Como o complexo conjugado do vetor ψ^α é $\bar{\psi}_\alpha$, sua norma então é dada por $\bar{\psi}_\alpha\psi^\alpha = \langle\bar{\psi}|\psi\rangle$.

Um operador em mecânica quântica pode ser representado como um mapa linear $T^\alpha_\beta : \mathcal{H}^+ \rightarrow \mathcal{H}^+$ dado por $\psi^\beta \mapsto T^\alpha_\beta \psi^\beta = \eta^\alpha$, para o qual corresponde um complexo conjugado $\overline{T^\alpha_\beta} = \bar{T}_\alpha^\beta$. Desta forma, se T é um operador Hermitiano, nos temos $T^\alpha_\beta = \bar{T}_\beta^\alpha$ e segue que o valor esperado $\langle T \rangle$ no estado ψ é dado por

$$\frac{\langle\bar{\psi}|T|\psi\rangle}{\langle\bar{\psi}|\psi\rangle} = \frac{\bar{\psi}_\alpha T^\alpha_\beta \psi^\beta}{\psi^\alpha \bar{\psi}_\alpha} = \frac{\bar{T}_\beta^\alpha \bar{\psi}_\alpha \psi^\beta}{\psi^\alpha \bar{\psi}_\alpha},$$

e a variância de T é

$$\text{Var}_\psi[T] = \frac{\bar{\psi}_\alpha \tilde{T}^\alpha_\gamma \tilde{T}^\gamma_\beta \psi^\beta}{\psi^\alpha \bar{\psi}_\alpha},$$

onde $\tilde{T}^\alpha_\beta = T^\alpha_\beta - \langle T \rangle \delta^\alpha_\beta$. Claramente $\langle \tilde{T} \rangle = 0$.

Agora vamos desenvolver mais a condição de hermiticidade. Tendo a decomposição de $\xi^a = (\xi^\alpha, \bar{\xi}_\alpha)$ em mente, podemos representar um tensor T_{ab} geral matricialmente da forma

$$T_{ab} = \begin{pmatrix} A_{\alpha\beta} & B_\alpha^\beta \\ C^\alpha_\beta & D^{\alpha\beta} \end{pmatrix}.$$

Similarmente, a estrutura complexa J^a_b pode ser escrita como

$$J^a_b = \begin{pmatrix} i\delta^\alpha_\beta & 0 \\ 0 & -i\delta_\alpha^\beta \end{pmatrix},$$

assim, obtemos $\zeta^a \equiv J^a_b \xi^b = (i\xi^\alpha, -i\bar{\xi}_\alpha)$. Em outras palavras, o efeito do tensor J^a_b é multiplicar o ket por i e o bra por $-i$ do estado em questão. Além disso, podemos verificar que

$$J^c_a J^d_b T_{cd} = \begin{pmatrix} -A_{\alpha\beta} & B_\alpha^\beta \\ C^\alpha_\beta & -D^{\alpha\beta} \end{pmatrix}.$$

Para a estrutura simplética Ω_{ab} temos

$$\Omega_{ab} = \begin{pmatrix} 0 & -i\delta_\alpha^\beta \\ +i\delta^\alpha_\beta & 0 \end{pmatrix}.$$

Já no caso da métrica g_{ab} ficamos com

$$g_{ab} = \begin{pmatrix} 0 & \delta_\alpha^\beta \\ \delta^\alpha_\beta & 0 \end{pmatrix}.$$

Agora podemos exigir algumas propriedades do tensor T_{ab} :

- i) Simétrico: $T_{ab} = T_{ba}$. Implicando que, $A_{\alpha\beta} = A_{(\alpha\beta)}$, $D^{\alpha\beta} = D^{(\alpha\beta)}$ e $B_\alpha^\beta = C^\beta_\alpha$.

ii) Hermitiano: $T_{ab} = J_a^c J_b^d T_{cd}$. Implicando $A_{\alpha\beta} = 0$ e $D^{\alpha\beta} = 0$.

iii) Real: $T_{ab} = \bar{T}_{ab}$. Implicando $B_\alpha^\beta = \bar{C}^\beta_\alpha$.

Portanto, um tensor T_{ab} real, hermitiano e simétrico, lembrando da condição de hermiticidade $C_\alpha^\beta = \bar{C}^\beta_\alpha$, tem representação matricial dada por

$$T_{ab} = \begin{pmatrix} 0 & T_\alpha^\beta \\ T_\beta^\alpha & 0 \end{pmatrix}.$$

Segue que a forma quadrática $T_{ab}\xi^a\eta^b$ para um tensor T como acima é dada por $T_{ab}\xi^a\eta^b = T_\alpha^\beta \xi^\alpha \bar{\eta}_\beta + T_\beta^\alpha \bar{\xi}_\alpha \eta^\beta$. Das matrizes g_{ab} e Ω_{ab} segue, respectivamente, que $g_{ab}\xi^a\eta^b = \xi^\alpha \bar{\eta}_\alpha + \bar{\xi}_\alpha \eta^\alpha$ e $\Omega_{ab}\xi^a\eta^b = -i\xi^\alpha \bar{\eta}_\alpha + i\bar{\xi}_\alpha \eta^\alpha$. Destas últimas derivamos que $(g_{ab} + i\Omega_{ab})\xi^a\eta^b = 2\xi^\alpha \bar{\eta}_\alpha$ e portanto

$$\xi^\alpha \bar{\eta}_\alpha = \frac{1}{2}(g_{ab} + i\Omega_{ab})\xi^a\eta^b = \frac{1}{2}\eta^a(g_{ab} - i\Omega_{ab})\xi^b,$$

que é consistente com o produto de Dirac em (5.12), visto que o tensor Ω_{ab} é antissimétrico. Com esses resultados, vemos que a reformulação com base em termos reais do formalismo complexo padrão da teoria quântica se dá de maneira direta e sistemática, não alterando os resultados ou o fundamento físico da teoria.

5.2.2 Dinâmica do tipo Schrödinger

Sejam X_{ab} e Y_{ab} um par de operadores simétricos. Seu *produto antissimétrico* é definido como $A_{ab} \equiv X_a^c Y_{cb} - X_b^c Y_{ca}$ que evidentemente é um tensor antissimétrico, portanto não pode representar uma variável aleatória. Mesmo assim, no caso de operadores hermitianos existe um isomorfismo entre o espaço dos tensores simétricos X_{ab} satisfazendo $J_a^c J_b^d X_{cd} = X_{ab}$ e o espaço dos tensores antissimétricos Λ_{ab} satisfazendo $J_a^c J_b^d \Lambda_{cd} = \Lambda_{ab}$. O mapa em questão é dado pela contração com a estrutura complexa J_b^a .

Proposição 8. Se X_{ab} é um tensor simétrico e hermitiano, então $\Lambda_{ab} = X_{ac} J_b^c$ é automaticamente antissimétrico e hermitiano.

Prova.

$$\begin{aligned} J_a^d J_e^b \Lambda_{ab} &= J_a^d J_e^b X_{ac} J_b^c = J_a^d J_b^c X_{ac} J_e^b = X_{db} J_e^b = \Lambda_{de} \\ &= J_b^c J_e^b X_{ac} J_a^d = -X_{ea} J_d^a = -\Lambda_{ed}. \end{aligned}$$

Claramente vemos que Λ_{ab} é hermitiano e antissimétrico se X_{ab} é simétrico e hermitiano. Por outro lado, pelo mesmo raciocínio da prova, temos que se Λ_{ab} é hermitiano e antissimétrico, então $\Lambda_{ac} J_b^c$ é automaticamente simétrico e hermitiano. $\square$

Assim sendo, esta propriedade nos indica que para formar em *comutador* entre dois tensores simétricos hermitianos temos que pegar seu produto antissimétrico e operarmos a estrutura complexa sobre ele, obtendo um tensor simétrico hermitiano.

Teorema 12. O comutador de um par de operadores simétricos hermitianos X_{ab} e Y_{ab} é dado pelo operador simétrico hermitiano $Z_{ab} = (X_{ac}Y_{bd} - Y_{ac}X_{bd})\Omega^{cd}$.

Prova.

$$\begin{aligned} Z_{ab} &= A_{ad}J_b^d = (X_a^b Y_{bd} - X_d^c Y_{ca})J_b^d \\ &= (g^{cb}X_{ac}Y_{bd} - g^{cb}X_{db}Y_{ca})J_b^d \\ &= (X_{ac}Y_{bd} - X_{bd}Y_{ac})g^{cb}J_b^d \\ &= (X_{ac}Y_{bd} - Y_{ac}X_{bd})\Omega^{cd} \end{aligned}$$

□

Agora, suponha que o Hamiltoniano de um sistema quântico, descrito pelo estado uniparamétrico $\xi^a(t)$, é representado por um operador simétrico hermitiano H_{ab} , então para uma *dinâmica do tipo Schrödinger* temos

$$\dot{\xi}^a = -J_b^a H_c^b \xi^c.$$

Expressando a dinâmica em termos de sua parte positiva e negativa, recuperamos a forma convencional da equação de Schrödinger junto com sua complexa conjugada, onde

$$\begin{aligned} \dot{\xi}_{\pm}^a &= \frac{1}{2}(\dot{\xi}^a \mp iJ_b^a \dot{\xi}^b) = \frac{1}{2}(\dot{\xi}^a \mp (-i)J_b^a J_c^b H_d^c \xi^d) \\ &= \frac{1}{2}(-H_d^a J_e^d \xi^e \mp iH_d^a \xi^d) = \frac{\mp i}{2}H_d^a (\mp iJ_e^d \xi^e + \xi^d) = \mp iH_d^a \xi_{\pm}^d, \end{aligned}$$

que na notação de Dirac tem parte positiva $i\partial_t |\xi\rangle = H |\xi\rangle$ e negativa $-i\partial_t \langle\xi| = H \langle\xi|$.

Pela Prop. 8, $H_{ac}J_b^c$ é automaticamente antissimétrico hermitiano, somando este fato com a dinâmica de Schrödinger temos que $\dot{\xi}_a \xi^a = 0$, como esperado, já que a equação de Schrödinger deve respeitar a normalização $g_{ab}\xi^a \xi^b$. A inserção da noção de uma fase global pode ser incorporada por uma modificação do Hamiltoniano de acordo com transformação $H_b^a \rightarrow \tilde{H}_b^a = H_b^a + \varphi\delta_b^a$. Agora podemos tirar vantagem disto tendo em mente o seguinte resultado.

Proposição 9. Existe apenas uma escolha de fase tal que o vetor tangente $\dot{\xi}^a$ da trajetória dinâmica é ortogonal para todo ponto com a direção $\zeta^a \equiv J_b^a \xi^b$. Esta mesma escolha de φ minimiza a informação de Fisher $\mathcal{G}$ e é dada por

$$\varphi = \frac{-H_{ab}\xi^a \xi^b}{\xi^c \xi_c} \quad (5.13)$$

Prova. Primeiro vamos mostrar qual o valor de φ tal que $\dot{\xi}^a \zeta_a = 0$ para todo ponto. A expressão

$$\dot{\xi}^a \zeta_a = -J_b^a J_a^d \tilde{H}_c^b \xi^c \xi_d = \delta_b^d (H_c^b + \varphi\delta_c^b) \xi^c \xi_d = 0,$$

resulta que temos de escolher o valor da fase que torna $H_c^d \xi^c \xi_d + \varphi \xi^c \xi_c$ nulo. Portanto, o cenário onde o vetor tangente $\dot{\xi}^a$ é sempre ortogonal a direção ζ_a , ou seja $\dot{\xi}^a \zeta_a = 0$, implica que $\varphi = -H_{ab} \xi^a \xi^b / \xi^c \xi_c$.

Agora vamos mostrar que o mesmo valor da fase extremiza (minimiza) a informação de Fisher.

$$\begin{aligned} \mathcal{G} &= 4g_{ab} \dot{\xi}^a \dot{\xi}^b = 4g_{ab} [J_c^a (H_c^d + \varphi \delta_c^d) \xi^d] [J_f^b (H_f^e + \varphi \delta_f^e) \xi^e] \\ &= 4g_{cf} [(H_c^d H_f^e + \varphi \delta_f^e H_c^d + \varphi \delta_c^d H_f^e + \varphi^2 \delta_c^d \delta_f^e) \xi^d \xi^e] \\ &= 4(H_{df} H_f^e + 2\varphi H_{de} + \varphi^2 g_{de}) \xi^d \xi^e \\ \min_{\varphi}(\mathcal{G}) \rightarrow \left. \frac{d\mathcal{G}}{d\varphi} \right|_{\tilde{\varphi}} &= 0 = 2H_{de} \xi^d \xi^e + 2\varphi \xi^d \xi_d \rightarrow \tilde{\varphi} = -H_{ab} \xi^a \xi^b / \xi^c \xi_c, \end{aligned}$$

Logo, o valor que extremiza $\mathcal{G}$ é exatamente o mesmo que faz $\dot{\xi}^a \zeta_a = 0$. $\square$

Fisicamente, a escolha de fase em (5.13) é aquela que leva a um ajuste da média do Hamiltoniano, onde $\tilde{H}_{ab} \xi^a \xi^b = 0$. Fixa esta escolha de fase, a *equação de Schrödinger modificada* fica

$$\dot{\xi}^a = -J_b^a \tilde{H}_c^b \xi^c$$

onde

$$\tilde{H}_{ab} = H_{ab} - \left(\frac{H_{cd} \xi^c \xi^d}{\xi^e \xi_e} \right) g_{ab},$$

representa agora o desvio da média do Hamiltoniano. Note que para o estado ζ^a , a equação dinâmica se torna $\dot{\zeta}^a = -J_b^a \tilde{H}_c^b \zeta^c$, já que H_{ab} é hermitiano, assim ζ^a também satisfaz a equação de Schrödinger. Pensando em termos do espaço projetivo complexo formado pela projetivização de $\mathcal{H}^+$ como sendo o verdadeiro espaço dos estados puros, temos então que a essência da Prop. 9 é que existe uma única fase tal que o $\dot{\xi}^a$ é ortogonal em todo ponto ξ^a e ζ^a .

No ensino padrão de mecânica quântica nos deparamos com a equação de Schrödinger independente do tempo. Tendo em vista a equação de Schrödinger modificada temos, para a parte positiva em termos da notação de Dirac, que $i\partial_t |\xi\rangle = (H - \langle H \rangle) |\xi\rangle$. Consequentemente se o estado é independente do tempo, recuperamos a equação de Schrödinger independente do tempo $H |\xi\rangle = E |\xi\rangle$. Enquanto em termos gerais a teoria é independente de uma escolha específica de fase, parece que existe uma única escolha de fase que faz com que tudo se ajuste bem de um ponto de vista físico. O interessante é que somos levados ao mesmo resultado a partir de considerações puramente estatísticas e matemáticas. (vide apêndices B e D)

Agora suponha B_{ab} operador simétrico hermitiano tal que $B(t) \equiv E_{\xi(t)}[B]$ onde

$$E_{\xi(t)}[B] = \frac{B_{ab} \xi^a(t) \xi^b(t)}{g_{cd} \xi^c(t) \xi^d(t)}, \quad (5.14)$$

onde $\xi^a(t)$ satisfaz a equação de Schrödinger modificada. Desta forma, $B(t)$ representa o valor esperado de B_{ab} ao longo da trajetória dada. A derivada de $B(t)$ em relação ao parâmetro t é dada por $\dot{B}(t) = 2B_{ab}\dot{\xi}^a\xi^b/g_{ef}\xi^e\xi^f$. Olhando para o numerador da derivada

$$\begin{aligned} 2B_{ab}\dot{\xi}^a\xi^b &= -[B_{ab}\Omega^{ac}\tilde{H}_{cd} - \tilde{H}_{ab}\Omega^{ac}B_{cd}]\xi^d\xi^b \\ &= -\Omega^{ac}([B_{ab}H_{cd} - H_{ab}B_{cd}] + \langle H \rangle [B_{ab}g_{cd} - g_{ab}B_{cd}])\xi^d\xi^b \\ &= -(\Omega^{ac}[B_{ab}H_{cd} - H_{ab}B_{cd}] + \langle H \rangle [B_{ab}J^a_d - J^c_b B_{dc}])\xi^d\xi^b = C_{db}\xi^d\xi^b, \end{aligned}$$

onde $C_{db} \equiv -\Omega^{ac}(B_{ab}H_{cd} - H_{ab}B_{cd})$ é o comutador entre B_{ab} e H_{ab} , como visto no Teo. 12. Segue que

$$\frac{dB(t)}{dt} = \frac{C_{ab}\xi^a\xi^b}{g_{cd}\xi^c\xi^d}.$$

A relação logo acima é a versão real de uma bem familiar: $d\langle B \rangle / dt = i\langle [H, B] \rangle$, no caso do operador B_{ab} não depender explicitamente de t . Note que o uso da equação modificada de Schrödinger não afeta este resultado.

Caso P_{ab} e Q_{ab} sejam operadores simétricos hermitianos satisfazendo a relação de comutação

$$(P_{ac}Q_{bd} - Q_{ac}P_{bd})\Omega^{cd} = g_{ab}, \quad (5.15)$$

então dizemos que P_{ab} é *canonicamente conjugado* a Q_{ab} , e nos referimos à relação acima como uma comutação canônica do tipo Heisenberg. Esta relação pode ser aplicada, por exemplo, quando P_{ab} e Q_{ab} são os operadores autoadjuntos momento e posição de um sistema quântico.

5.2.3 Estimativa quântica

Considere uma família de vetores de estado normalizados $\xi^a(t)$, parametrizados apenas pelo tempo t satisfazendo a equação de Schrödinger modificada. A curva descrita por $\xi^a(t)$ vive em $\mathcal{P} \in \mathcal{H}$. Agora estamos em condição de estudar o problema de estimar o parâmetro t usando as técnicas geométricas desenvolvidas ao longo da Sec. 5.1. Seja T_{ab} um estimador não viesado para o parâmetro t . Deste modo T_{ab} é um operador real, simétrico e hermitiano satisfazendo

$$\frac{T_{ab}\xi^a(t)\xi^b(t)}{g_{cd}\xi^c(t)\xi^d(t)} = t, \quad (5.16)$$

para um sistema quântico que está no estado $\xi^a(t)$.

A *ideia* é aplicar as cotas generalizadas de Bhattacharyya estabelecidas na Sec. 5.1.4 no problema de estimativa quântica, considerando a possibilidade de estabelecer correções nas relações de incerteza do tipo Heisenberg $\Delta H^2 \Delta T^2 \geq 1/4$ no caso de variáveis canonicamente conjugadas. Lembrando o conteúdo geométrico das cotas generalizadas de Bhattacharyya, dado o vetor normal $\nabla^a t$ as superfícies que folheiam $\mathcal{P}$, para cada t constante, podemos escolher um conjunto de vetores ortogonais em $\mathcal{H}$ e expressar o

comprimento deste vetor em termos destas componentes ortogonais. Usando a Prop. 6 juntamente com as cotas generalizadas de Bhattacharyya da Prop. 7, podemos formular as correções na cota para a variância do estimador T_{ab} .

No cenário clássico de estimativa de parâmetros construímos a base (5.9), dada por $\hat{\xi}^{(k)a}$ ($k \in \mathbb{N}^\times$), representando a k -ésima derivada do estado ξ^a projetada ortogonalmente às derivadas de ordem inferior a k que será denotado *sistema clássico ortogonal de vetores*. Em contrapartida, no caso da mecânica quântica de estado puro podemos construir um conjunto de vetores ortogonais mais rico, já que o tensor estrutura complexo pode ser colocado na cena. Sabemos que $\zeta^a = J^a_b \xi^b$ é ortogonal a ξ^a , pela dinâmica da Schrödinger modificada $\dot{\zeta}^a$ é ortogonal a ξ^a , e assim por diante, portanto podemos construir um conjunto ortogonal de vetores a partir de ζ^a e ξ^a e suas derivadas de ordem superior. Este conjunto será denotado por *sistema quântico ortogonal de vetores*, onde $\check{\xi}^{(k)a}$ e $\check{\zeta}^{(k)a}$ ($k \in \mathbb{N}^\times$) satisfazem $\check{\xi}_a^{(j)} \check{\xi}^{(k)a} = 0$, $\check{\zeta}_a^{(j)} \check{\zeta}^{(k)a} = 0$ e $\check{\xi}_a^{(j)} \check{\zeta}^{(k)a} = 0$ para $j \neq k$. Note que nada prioriza a princípio a escolha de um sistema pelo outro, ambos são bases ortogonais válidas que só vão produzir correções distintas para a cota inferior, afinal bases diferentes produzem projeções diferentes.

Antes de partirmos para ordens superiores, que serão abordadas na próxima seção, vamos focar nos termos de ordem $k = 1$ para o sistema quântico ortogonal de vetores. Neste caso a cota da variância é dada por

$$\text{Var}_\xi[T] \geq \frac{(\dot{\xi}^a \nabla_a t)^2}{4 \dot{\xi}^b \dot{\xi}_b} + \frac{(\dot{\zeta}^a \nabla_a t)^2}{4 \dot{\zeta}^b \dot{\zeta}_b}. \quad (5.17)$$

Proposição 10. Seja $\xi^a(t)$ satisfazendo a equação modificada de Schrödinger, e seja $\zeta^a = J^a_b \xi^b$. Dado que T_{ab} é o estimador não viesado do parâmetro t , então $\dot{\xi}^a \nabla_a t = 1$ e $\dot{\zeta}^a \nabla_a t = 2 \text{Cov}_\xi[H, T]$, onde $\text{Cov}_\xi[H, T]$ é a covariância entre os operadores H_{ab} e T_{ab} no estado ξ^a .

Prova. Ao longo dessa demonstração tenha em mente o resultado da Prop. 8. Iniciamos derivando a expressão (5.16) com respeito ao parâmetro t obtendo

$$\begin{aligned} \frac{T_{ab}[\dot{\xi}^a(t) \xi^b(t) + \xi^a(t) \dot{\xi}^b(t)]}{g_{cd} \xi^c(t) \xi^d(t)} &= \frac{2T_{ab} \xi^a \dot{\xi}^b}{g_{cd} \xi^c \xi^d} = \frac{-2T_{ab} J^b_e \tilde{H}^e_f \xi^f \xi^a}{g_{cd} \xi^c \xi^d} \\ &= \frac{-2T_{ab} J^b_e g^{eh} H_{hf} \xi^f \xi^a}{g_{cd} \xi^c \xi^d} = -\frac{2T_{ab} H_{hf} \Omega^{bh} \xi^f \xi^a}{g_{cd} \xi^c \xi^d} = 1 \end{aligned}$$

para qualquer vetor de estado em uma trajetória específica $\xi^a(t)$. Por outro lado, $\dot{\xi}^a \nabla_a t = -J^a_b \tilde{H}^b_c \xi^c \nabla_a t$. Da expressão (5.8) e o resultado logo acima obtemos

$$\dot{\xi}^a \nabla_a t = -J^a_b \tilde{H}^b_c \xi^c \frac{2\tilde{T}_{ad} \xi^d}{g_{ef} \xi^e \xi^f} = \frac{-2\tilde{T}_{ad} \tilde{H}_{hc} \Omega^{ah} \xi^c \xi^d}{g_{ef} \xi^e \xi^f} = -\frac{2T_{ad} H_{hc} \Omega^{ah} \xi^c \xi^d}{g_{ef} \xi^e \xi^f} = 1,$$

como foi requerido.

No que segue, observamos que

$$\begin{aligned}\dot{\zeta}^a &= J^a_b \dot{\xi}^b = J^a_b (-J^b_c \tilde{H}^c_d \xi^d) = J^a_b (-\tilde{H}^b_c J^c_d \xi^d) = -J^a_b \tilde{H}^b_c \xi^c \\ &= -\tilde{H}^a_b J^b_c J^c_e \xi^e = +\tilde{H}^a_b \xi^b,\end{aligned}$$

desta maneira

$$\dot{\zeta}^a \nabla_{at} = \tilde{H}^a_b \xi^b \frac{2\tilde{T}_{ad}\xi^d}{g_{ef}\xi^e\xi^f} = \frac{2\tilde{H}_{bc}\tilde{T}^c_d \xi^b \xi^d}{g_{ef}\xi^e\xi^f} = 2\text{Cov}_\xi[H, T].$$

□

Atente ao fato que a informação de Fisher quando o valor da fase corresponde àquele visto na Eq. (5.13) faz $\mathcal{G} = 4[H^c_a H_{bc} \xi^a \xi^b - (H_{ab} \xi^a \xi^b)^2] = 4\text{Var}_\xi[H]$. Também temos que $4\dot{\zeta}^a \dot{\zeta}_a = 4\tilde{H}_{ab} \tilde{H}^b_c \xi^a \xi^c = 4\text{Var}_\xi[H]$.

Se redefinirmos a notação das variâncias de forma que agora $\Delta_\xi H^2 := \text{Var}_\xi[H]$ e $\Delta_\xi T^2 := \text{Var}_\xi[T]$ a correção de primeira ordem (5.17) da cota para a variância quântica é dada por

$$\Delta_\xi T^2 \Delta_\xi H^2 \geq \frac{1}{4} + \text{Cov}_\xi[H, T]. \quad (5.18)$$

Em particular, o segundo termo a direita da expressão acima pode ser reescrito utilizando o anticomutador via relação $\text{Cov}_\xi[H, T] = (1/2)E_\xi[\tilde{H}\tilde{T} + \tilde{T}\tilde{H}]$, onde $\tilde{H} = H - E_\xi[H]$ e $\tilde{T} = T - E_\xi[T]$. Como o termo da covariância depende explicitamente da escolha do estimador, vamos omiti-lo ficando apenas com o primeiro termo, temos então o *CRB quântico*

$$\Delta_\xi T^2 \Delta_\xi H^2 \geq \frac{1}{4}. \quad (5.19)$$

5.2.4 Correções de ordem superior na cota da variância quântica

A cota dada no CRB quântico é a mesma que a relação de incerteza de Heisenberg. Entretanto, na Sec. 5.1.3 vimos que para saturarmos a cota é necessário que o estado ξ^a satisfaça a equação diferencial (5.6), onde $\tilde{T}_{ab}$ é um tensor *simétrico*. Por outro lado, a forma $\Omega_{ac}\tilde{H}^c_b$ definida na equação de Schrödinger modificada é *antissimétrica*, graças a condição de hermiticidade e simetria do Hamiltoniano. Se $\xi^a(t)$ satisfaz a equação modificada temos que $\xi^a(t) = \exp[t\Omega_{ac}\tilde{H}^c_b]\xi^b(0)$, onde $\xi^a(0)$ é um vetor de estado normalizado em $t = 0$, logo, dinâmica do tipo Schrödinger não pode gerar famílias exponenciais de distribuições em t . Em particular, no caso da estimativa uniparamétrica no contexto quântico puro, relações de incerteza do tipo Heisenberg tem sua cota mínima inatingível. Do ponto de vista experimental, agora se torna muito relevante saber sobre as correções de ordem superior da cota quântica.

Como já destacamos ao analisar a relação (5.18), o primeiro termo à direita não depende da escolha específica do estimador, já o segundo envolve a covariância entre

H_{ab} e T_{ab} dependendo da escolha de T_{ab} . Desta forma jogamos fora o segundo termo, pois estávamos querendo considerar relações de incerteza, claro que a cota geral é menos *tight** que aquela vista no CRB quântico. Note que não fizemos nenhuma consideração na derivação de (5.18) do fato que T_{ab} ser canonicamente conjugado a H_{ab} , apenas assumimos que T_{ab} é um estimador de t para uma dada trajetória $\xi^a(t)$ de acordo com a expressão (5.16). Pedir conjugação canônica torna possível derivar cotas independentes da escolha de um estimador específico. Assim sendo, focaremos nas ordens de correção superiores do tipo Bhattacharyya para o CRB encontrando algumas correções que dependem da escolha de T_{ab} e outras que não, os termos que não dependem da escolha específica do estimador vão contribuir para um conjunto de *relações generalizadas de Heisenberg* para a inferência estatística quântica.

Contudo antes de iniciar os cálculos, vamos apresentar alguns resultados auxiliares. Seja $\xi^a(t)$ satisfazendo a Equação de Schrödinger modificada.

Proposição 11. $g_{ab}\xi^{(n)a}\xi^{(n)b} = \langle \tilde{H}^{2n} \rangle$, onde $\langle \tilde{H}^r \rangle$ corresponde ao r -ésimo momento do Hamiltoniano sobre sua média, sendo $\xi^{(n)a}$ a n -ésima derivada de $\xi^a(t)$ com respeito ao tempo t .

Prova. Como $\tilde{H}_{ab}$ é simétrico hermitiano, é fácil ver que

$$\xi^{(n)a} = (-1)^n J_{i_1}^a J_{i_2}^{i_1} \dots J_{i_n}^{i_{n-1}} \tilde{H}_{i_{n+1}}^{i_n} \dots \tilde{H}_{i_{2n}}^{i_{2n-1}} \xi^{i_{2n}}.$$

Inserindo notação

$$\begin{aligned} \text{mod}_{(J_b^a)}[n] &\Rightarrow n = \bar{1} \rightarrow J_b^a \\ n = \bar{2} &\rightarrow -\delta_b^a \\ n = \bar{3} &\rightarrow -J_b^a \\ n = \bar{4} &\rightarrow \delta_b^a, \end{aligned}$$

onde $\bar{n}$ representa a classe de números múltiplos inteiros de $n \in \mathbb{N}^\times$.

Podemos agora reescrever $\xi^{(n)a}$ como

$$\xi^{(n)a} = (-1)^{n \text{mod}_{(J_{i_1}^a)}[n]} \cdot \tilde{H}_{i_2}^{i_1} \dots \tilde{H}_{i_n}^{i_{n-1}} \xi^{i_n}.$$

Basta calcularmos então a norma ao quadrado de $\xi^{(n)a}$

$$\begin{aligned} g_{ab}\xi^{(n)a}\xi^{(n)b} &= (-1)^{2n} g_{ab} \text{mod}_{(J_{i_1}^a)}[n] \cdot \text{mod}_{(J_{i_1}^b)}[n] \cdot \tilde{H}_{i_2}^{i_1} \dots \tilde{H}_{i_n}^{i_{n-1}} \tilde{H}_{j_2}^{j_1} \dots \tilde{H}_{j_n}^{j_{n-1}} \xi^{i_n} \xi^{j_n} \\ &= g_{i_1 j_1} \tilde{H}_{i_2}^{i_1} \dots \tilde{H}_{i_n}^{i_{n-1}} \tilde{H}_{j_2}^{j_1} \dots \tilde{H}_{j_n}^{j_{n-1}} \xi^{i_n} \xi^{j_n} = \langle \tilde{H}^{2n} \rangle \end{aligned}$$

□

*A palavra *tight*, traduzida de forma literal do inglês, significa “justo”, “apertado”. Uma cota mais *tight* fornece uma estimativa melhor quando saturada. Na falta de uma expressão em português que expresse precisamente esta ideia, vamos utilizar no texto o termo na sua notação da literatura especializada, ou seja, o termo *tight*.

Proposição 12. Para um estado do tipo Schrödinger ξ^a , os momentos pares do hamiltoniano sobre sua média são independentes do parâmetro t .

Prova. Derivando a norma ao quadrado de $\xi^{(n)a}$ apresentada na Prop. 11 temos

$$\begin{aligned} \frac{d}{dt} g_{ab} \xi^{(n)a} \xi^{(n)b} &= 2g_{ab} \xi^{(n)a} \dot{\xi}^{(n)b} = -2\xi^{(n)a} g_{ab} J^b_c \tilde{H}^c_d \xi^{(n)d} \\ &= -2g_{ab} \xi^{(n)a} \xi^{(n+1)b} = -2J_{ac} \tilde{H}^c_d \xi^{(n)a} \xi^{(n)d} = 0, \end{aligned}$$

onde $J_{ac} \tilde{H}^c_d$ é antissimétrico pelo Prop. 8.

Uma consequência direta é que para um n arbitrário

$$g_{ab} \xi^{(n)a} \xi^{(n+1)b} = 0,$$

um resultado muito útil na busca por correções de ordem superior independentes da escolha do estimador. $\square$

Proposição 13. Seja T_{ab} canonicamente conjugado a H_{ab} e estimador não viesado para o parâmetro t . Então

$$T_{ab} \xi^{(n)a} \xi^{(n)b} = t g_{ab} \xi^{(n)a} \xi^{(n)b} + \kappa,$$

onde κ é uma constante. Assim, para todo n , $\tilde{T}_{ab} \xi^{(n)a} \xi^{(n)b} = \kappa$ é uma constante de movimento ao longo da trajetória do tipo Schrödinger.

Prova. Se T_{ab} é canonicamente conjugado a H_{ab} , então pela relação (5.15)

$$(H_{ab} T_{cd} - T_{ab} H_{cd}) \Omega^{ac} = -2T_{ab} \Omega^{ac} H_{cd} = g_{bd}.$$

A equação de Schrödinger para $\xi^{(n)a}$ tem a forma $\dot{\xi}^{(n)a} = -J^a_b g^{bc} \tilde{H}_{cd} \xi^{(n)d} = -\Omega^{ac} \tilde{H}_{cd} \xi^{(n)d}$. Derivando $T_{ab} \xi^{(n)a} \xi^{(n)b}$ com respeito a t chegamos em

$$\frac{d}{dt} T_{ab} \xi^{(n)a} \xi^{(n)b} = 2T_{ab} \dot{\xi}^{(n)a} \xi^{(n)b} = -2T_{ab} \Omega^{ac} \tilde{H}_{cd} \xi^{(n)d} \xi^{(n)b} = -2T_{ab} \Omega^{ac} H_{cd} \xi^{(n)d} \xi^{(n)b}.$$

Unindo estes dois resultados acima temos

$$-2T_{ab} \Omega^{ac} H_{cd} \xi^{(n)d} \xi^{(n)b} = g_{bd} \xi^{(n)d} \xi^{(n)b} = \langle \tilde{H}^{2n} \rangle,$$

que segundo a Prop. 12 é independente de t .

Integrando a expressão acima obtemos o resultado requerido

$$\begin{aligned} T_{ab} \xi^{(n)a} \xi^{(n)b} &= t g_{ab} \xi^{(n)a} \xi^{(n)b} + \kappa \\ \tilde{T}_{ab} \xi^{(n)a} \xi^{(n)b} &= \kappa, \end{aligned}$$

onde κ é uma constante independente de t e $\tilde{T}_{ab} \xi^{(n)a} \xi^{(n)b}$ é uma constante de movimento. $\square$

Proposição 14. Seja T_{ab} canonicamente conjugado a H_{ab} e estimador não viesado para o parâmetro t . Então para inteiros ímpares n , com $m = (n - 1)/2$ vale

$$2T_{ab}\xi^{(n)a}\xi^b = (-1)^m n g_{ab} \xi^{(m)a} \xi^{(m)b}$$

Esquema de prova. • $n = 1$: seque que de $T_{ab}\xi^a\xi^b = t \xrightarrow{d/dt} 2T_{ab}\dot{\xi}^a\xi^b = 1$.

• $n = 3$: diferenciando $T_{ab}\xi^a\xi^b = t$ mais duas vezes temos $T_{ab}\ddot{\xi}^a\xi^b + 3T_{ab}\dot{\xi}^a\dot{\xi}^b = 0$. Tendo em mente a Prop. 12 e a Prop. 13 deduzimos $T_{ab}\dot{\xi}^a\dot{\xi}^b = t\dot{\xi}^a\ddot{\xi}_a + \kappa \xrightarrow{d/dt} 2T_{ab}\ddot{\xi}^a\xi^b = \dot{\xi}^a\dot{\xi}_a$. Dados esses resultados temos o resultado desejado: $2T_{ab}\ddot{\xi}^a\xi^b = -3\dot{\xi}^a\dot{\xi}_a$.

• $n = 5$: diferenciando $T_{ab}\xi^a\xi^b = t$ cinco vezes temos $T_{ab}(\xi^{(5)a}\xi^b + 5\xi^{(4)a}\dot{\xi}^b + 10\dot{\xi}^a\ddot{\xi}^b) = 0$. Tendo em mente a Prop. 12 e a Prop. 13 deduzimos $T_{ab}\ddot{\xi}^a\ddot{\xi}^b = t\ddot{\xi}^a\ddot{\xi}_a \xrightarrow{d/dt} 2T_{ab}\ddot{\xi}^a\dot{\xi}^b = \ddot{\xi}^a\dot{\xi}_a$; derivando $2T_{ab}\ddot{\xi}^a\dot{\xi}^b = \dot{\xi}^a\dot{\xi}_a$ duas vezes chegamos em $T_{ab}\xi^{(4)a}\dot{\xi}^b = -3T_{ab}\ddot{\xi}^a\dot{\xi}^b$. Dados esses resultados temos o resultado desejado: $2T_{ab}\xi^{(5)a}\xi^b = 5\ddot{\xi}^a\dot{\xi}_a$. Formulas de ordens superiores podem ser deduzidas da mesma forma. $\square$

Em posse desses resultados estamos prontos para deduzir algumas correções de ordem superior de relações de incerteza do tipo Heisenberg para observáveis canonicamente conjugados. O contexto é novamente o problema de estimativa do parâmetro t no caso de uma família de vetores de estado uniparamétrica $\xi^a(t)$ que obedece a equação de Schrödinger modificada, dado Hamiltoniano H_{ab} . O observável T_{ab} é o estimador não viesado do parâmetro t tomado canonicamente conjugado ao Hamiltoniano.

Consideraremos correções até a segunda ordem para a cota da variância de T_{ab} que emergem quando expandimos $\nabla^a t$ no sistema quântico ortogonal de vetores, dados por $\dot{\xi}^a$, $\check{\xi}^a$, $\check{\xi}^{(2)a}$ e $\check{\xi}^{(2)a}$. Usando as cotas generalizadas de Bhattacharyya temos

$$\text{Var}_\xi[T] \geq \frac{(\dot{\xi}^a \nabla_a t)^2}{4\dot{\xi}^b \dot{\xi}_b} + \frac{(\check{\xi}^a \nabla_a t)^2}{4\check{\xi}^b \check{\xi}_b} + \frac{(\check{\xi}^{(2)a} \nabla_a t)^2}{4\check{\xi}^{(2)b} \check{\xi}_b^{(2)}} + \frac{(\check{\xi}^{(2)a} \nabla_a t)^2}{4\check{\xi}^{(2)b} \check{\xi}_b^{(2)}}, \quad (5.20)$$

onde $\check{\xi}^{(2)a}$ e $\check{\xi}^{(2)a}$ são dados por

$$\check{\xi}^{(2)a} = \ddot{\xi}^a - \frac{(\ddot{\xi}^b \dot{\xi}_b)}{\dot{\xi}^c \dot{\xi}_c} \dot{\xi}^a - (\ddot{\xi}^b \xi_b) \xi^a \text{ e} \quad (5.21)$$

$$\check{\xi}^{(2)a} = \ddot{\xi}^a - \frac{(\ddot{\xi}^b \dot{\xi}_b)}{\dot{\xi}^c \dot{\xi}_c} \dot{\xi}^a - (\ddot{\xi}^b \xi_b) \xi^a. \quad (5.22)$$

Aparecem apenas três termos nas expressões acima, pois $\ddot{\xi}^a \dot{\xi}_a = \ddot{\xi}^a \xi_a = 0$ e $\ddot{\xi}^a \xi_a = \ddot{\xi}^a \xi_a = 0$

Pode ser mostrado, utilizando a definição da estrutura complexa e hermiticidade da métrica, que as normas dos vetores $\check{\xi}^{(2)a}$ e $\check{\xi}^{(2)a}$ são iguais e dadas por

$$\check{\xi}^{(2)a} \check{\xi}_a^{(2)} = \check{\xi}^{(2)a} \check{\xi}_a^{(2)} = \langle \tilde{H}^2 \rangle K_\xi^2,$$

onde K_ξ^2 é definido como

$$K_\xi^2 \equiv \frac{\langle \tilde{H}^4 \rangle}{\langle \tilde{H}^2 \rangle^2} - \frac{\langle \tilde{H}^3 \rangle^2}{\langle \tilde{H}^2 \rangle^3} - 1. \quad (5.23)$$

Os primeiros dois termos a direita na (5.20) já foram calculados em (5.18). Seguimos então com o cálculo dos termos de segunda ordem. Os denominadores que aparecem na segunda ordem são dados por (5.23). Focando nos numeradores, o termo envolvendo $\check{\xi}^{(2)a}$, usando (5.8) e (5.21), somado com o fato que $\check{\xi}^{(2)a}\xi_a = 0$, requer o cálculo da expressão

$$2 \left(\underbrace{\check{\xi}^a}_{(I)} - \underbrace{\frac{(\check{\xi}^b \dot{\xi}_b)}{\dot{\xi}^c \dot{\xi}_c} \dot{\xi}^a}_{(II)} - \underbrace{(\check{\xi}^b \xi_b) \xi^a}_{(III)} \right) T_{ad} \xi^d.$$

Nomeando, da esquerda para a direita, os termos que aparecem acima por (I), (II) e (III), pela Prop. 14 temos

$$\begin{aligned} 2T_{ab}\dot{\xi}^a\xi^b &= 1 \xrightarrow{d/dt} T_{ab}\ddot{\xi}^a\xi^b = -T_{ab}\dot{\xi}^a\dot{\xi}^b \\ g_{ab}\dot{\xi}^a\xi^b &= 0 \xrightarrow{d/dt} \check{\xi}^a\xi_a = -\dot{\xi}^a\dot{\xi}_a, \end{aligned}$$

assim

$$\begin{aligned} (I)+(III) &= T_{ab}\ddot{\xi}^a\xi^b - (\check{\xi}^c\xi_c)T_{ab}\dot{\xi}^a\xi^b \\ &= -T_{ab}\dot{\xi}^a\dot{\xi}^b + (\dot{\xi}^c\dot{\xi}_c)T_{ab}\dot{\xi}^a\xi^b = -\tilde{T}_{ab}\dot{\xi}^a\dot{\xi}^b \\ &= -J^a{}_c J^b{}_e \tilde{T}_{ab} \tilde{H}^c{}_d \tilde{H}^e{}_f \xi^d \xi^f = -\tilde{H}^c{}_d \tilde{T}_{ce} \tilde{H}^e{}_f \xi^d \xi^f. \end{aligned}$$

Entretanto para operadores canonicamente conjugados vale a identidade $\tilde{T}\tilde{H} + \tilde{H}\tilde{T} = 2\tilde{H}\tilde{T}\tilde{H}$, então podemos escrever $(I)+(III) = -\{\tilde{T}, \tilde{H}^2\}_\xi/2$.

Seguindo com o cálculo, chegamos para termo (2) em

$$\begin{aligned} T_{ab}\dot{\xi}^a\xi^b &= \tilde{T}_{ab}\dot{\xi}^a\xi^b = \tilde{T}_{ab}\tilde{H}^a{}_c \xi^c \xi^b = \{\tilde{T}, \tilde{H}\}_\xi/2 \\ \check{\xi}^a &= -\delta^a{}_c \tilde{H}^c{}_d \tilde{H}^d{}_e \xi^e = -\tilde{H}^a{}_d \tilde{H}^d{}_e \xi^e \\ \check{\xi}^a \dot{\xi}_a &= -\tilde{H}^a{}_b \tilde{H}^b{}_c \tilde{H}_{ad} \xi^c \xi^d = -\langle \tilde{H}^3 \rangle \\ \dot{\xi}^a \dot{\xi}_a &= \tilde{H}^a{}_b \tilde{H}_{ac} \xi^b \xi^c = \langle \tilde{H}^2 \rangle \\ (II) &= \frac{\langle \tilde{H}^3 \rangle \{\tilde{T}, \tilde{H}\}_\xi}{2 \langle \tilde{H}^2 \rangle}. \end{aligned}$$

Combinando os resultados acima obtemos

$$\frac{(\check{\xi}^{(2)a} \nabla_a t)^2}{4 \check{\xi}^{(2)a} \check{\xi}_a^{(2)}} = \frac{1}{4 \langle \tilde{H}^2 \rangle K_\xi^2} \left(\{\tilde{T}, \tilde{H}^2\}_\xi - \frac{\langle \tilde{H}^3 \rangle}{\langle \tilde{H}^2 \rangle} \{\tilde{T}, \tilde{H}\}_\xi \right)^2. \quad (5.24)$$

Passando para o termo envolvendo $\check{\xi}^{(2)a}$, queremos calcular o valor da expressão

$$2 \left(\check{\xi}^a - \frac{(\check{\xi}^b \dot{\xi}_b)}{\dot{\xi}^c \dot{\xi}_c} \dot{\xi}^a - (\check{\xi}^b \xi_b) \xi^a \right) T_{ad} \xi^d. \quad (5.25)$$

Pela Prop. 14 temos que $T_{ab}\dot{\xi}^a\xi^b = 1/2$. Ainda valem os seguintes

$$\begin{aligned}\ddot{\xi}^a &= -J^a_b \tilde{H}^b_c \tilde{H}^c_d \xi^d \\ T_{ae}\ddot{\xi}^a\xi^e &= -\xi^e T_{ea} J^a_b \tilde{H}^b_c \tilde{H}^c_d \xi^d = 0 \\ T_{ab}\zeta^a\xi^b &= T_{ab}J^b_c \xi^c \xi^b = 0 \\ \dot{\xi}^a\ddot{\xi}_a &= (J_{ba}J^a_e)\tilde{H}^e_f \tilde{H}^b_c \tilde{H}^c_d \xi^f \xi^d = -\langle \tilde{H}^3 \rangle \\ \dot{\xi}^a\dot{\xi}_a &= \langle \tilde{H}^2 \rangle.\end{aligned}$$

Aplicando estes resultados à expressão (5.25) obtemos:

$$\frac{(\check{\zeta}^{(2)a}\nabla_a t)^2}{4\check{\zeta}^{(2)a}\check{\zeta}_a^{(2)}} = \frac{1}{4\langle \tilde{H}^2 \rangle K_\xi^2} \frac{\langle \tilde{H}^3 \rangle^2}{\langle \tilde{H}^2 \rangle^3}. \quad (5.26)$$

A contribuição dada por (5.24) depende da escolha de um estimador específico T_{ab} , por isso vamos omitir este termo. Considerando o termo representado em (5.26) obtemos uma correção para a cota inferior da variância do estimador T_{ab} que pode ser vista como uma forma generalizada da relação de incerteza de Heisenberg.

Proposição 15. Se T_{ab} e H_{ab} são canonicamente conjugados, então a cota

$$\langle \tilde{T}^2 \rangle \langle \tilde{H}^2 \rangle \geq \frac{1}{4} \left(1 + \frac{\langle \tilde{H}^3 \rangle^2}{\langle \tilde{H}^2 \rangle^3 K_\xi^2} \right) = \frac{1}{4} \left(1 + \frac{\langle \tilde{H}^3 \rangle^2}{\langle \tilde{H}^4 \rangle \langle \tilde{H}^2 \rangle - \langle \tilde{H}^3 \rangle^2 - \langle \tilde{H}^2 \rangle^3} \right)$$

se aplica ao produto das suas variâncias no estado de Schrödinger $\xi^a(t)$.

O fato do denominador da ser positivo nesta correção pode ser verificado diretamente notando que ele é o módulo quadrado do estado $|\psi\rangle$ definido como

$$|\psi\rangle = \left(\tilde{H}^2 - \frac{\langle \tilde{H}^3 \rangle}{\langle \tilde{H}^2 \rangle} \tilde{H} - \langle \tilde{H}^2 \rangle \right) |\xi\rangle,$$

onde $\langle \xi | \xi \rangle = 1$.

A fim de ilustrar a generalidade do formalismo, podemos exibir uma outra cota para a variância, distinta da vista acima, ainda independente da escolha de um estimador T_{ab} do parâmetro t , que surge ao considerarmos desigualdades baseadas na expansão de $\nabla_a t$ no sistema clássico ortogonal de vetores (5.9) associados a $\xi^a(t)$. Esta nova cota pode ser derivada quando considerarmos o termo de terceira ordem, dado por $(\hat{\xi}^{(3)a}\nabla_a t)^2/4\hat{\xi}^{(3)a}\hat{\xi}_a^{(3)}$. O termo de segunda ordem, vindo de $\hat{\xi}^{(2)a}$, é descartado pois irá depender explicitamente de T_{ab} .

Segundo a Prop. 12, $\dot{\xi}^a\dot{\xi}_a$ e $\ddot{\xi}^a\ddot{\xi}_a$ são constantes temporais ao longo da trajetória quântica então $\xi^{(3)a}\ddot{\xi}_a = \ddot{\xi}^a\dot{\xi}_a = 0$. Como $\dot{\xi}^a\xi_a = 0$, segue que $\ddot{\xi}^a\xi_a = -\dot{\xi}^a\dot{\xi}_a$, então $\xi^{(3)a}\xi_a = 0$. Assim podemos escrever

$$\hat{\xi}^{(3)a} = \xi^{(3)a} - \frac{\dot{\xi}^b \xi_b^{(3)}}{\dot{\xi}^c \xi_c} \dot{\xi}^a.$$

Basta agora calcularmos a cota da variância vinda do primeiro e terceiro termo

$$\text{Var}_\xi[T] \geq \frac{(\dot{\xi}^a \nabla_a t)^2}{4\dot{\xi}^b \dot{\xi}_b} + \frac{(\hat{\xi}^{(3)a} \nabla_a t)^2}{4\hat{\xi}^{(3)a} \hat{\xi}_a^{(3)}}.$$

Sabemos que o primeiro termo a direita da expressão acima resulta em $1/4 \langle \tilde{H}^2 \rangle$. Agora, de acordo com a Prop. 14, $\xi^{(3)a} \nabla_a t = -3 \langle \tilde{H}^2 \rangle$. Também é fácil mostrar que $\dot{\xi}^a \xi_a^{(3)} = -\langle \tilde{H}^4 \rangle$ e pela Prop. 11 $\xi^{(3)a} \xi_a^{(3)} = \langle \tilde{H}^6 \rangle$, dados estes resultados segue que

$$\hat{\xi}^{(3)a} \hat{\xi}_a^{(3)} = \langle \tilde{H}^6 \rangle - \frac{\langle \tilde{H}^4 \rangle^2}{\langle \tilde{H}^2 \rangle}.$$

Colocando estes ingredientes juntos, obtemos a seguinte correção para o CRB que só depende da trajetória $\xi^a(t)$ gerada por H_{ab} :

Proposição 16. Se T_{ab} e H_{ab} são canonicamente conjugados, então a cota

$$\langle \tilde{T}^2 \rangle \langle \tilde{H}^2 \rangle \geq \frac{1}{4} \left(1 + \frac{(\langle \tilde{H}^4 \rangle - 3 \langle \tilde{H}^2 \rangle^2)^2}{\langle \tilde{H}^6 \rangle \langle \tilde{H}^2 \rangle - \langle \tilde{H}^4 \rangle^2} \right). \quad (5.27)$$

se aplica ao produto das suas variâncias no estado de Schrödinger $\xi^a(t)$.

A positividade do denominador pode ser verificada diretamente notando que ele é o módulo quadrado do estado $|\psi\rangle$ definido como

$$|\psi\rangle = \left(\tilde{H}^3 - \frac{\langle \tilde{H}^4 \rangle}{\langle \tilde{H}^2 \rangle} \tilde{H} \right) |\xi\rangle \quad (5.28)$$

onde $\langle \xi | \xi \rangle = 1$.

5.3 A correção no caso misto

5.3.1 Adaptação da notação

Seja um espaço de Hilbert $\mathcal{H}$ com base ortonormal $\{e_i\}_{i=1,\dots,n}$. Seleccionamos o subconjunto $\mathcal{B}_{HS} \subset \mathcal{B} \subset M_n(\mathbb{C})$ de operadores que possuem norma de Hilbert-Schmidt finita, ou seja

$$\|A\|_{HS}^2 = \text{Tr}(A^\dagger A) = \sum_i \|Ae_i\|^2 < \infty \text{ tal que } A \in \mathcal{B},$$

onde $\|\cdot\|$ é norma de $\mathcal{H}$. denotamos tais operadores como sendo *operadores de Hilbert-Schmidt*. Note que $\dim \mathcal{B} = (\dim \mathcal{H})^2 = n^2$.

Para um par de operadores $A, B \in \mathcal{B}_{HS}$, dizemos que $\sum_i \langle Ae_i, Be_i \rangle$ é somável, onde esta soma define um produto interno neste subconjunto de $\mathcal{B}$

$$\langle A, B \rangle = \text{Tr}(A^\dagger B) = \sum_i \langle Ae_i, Be_i \rangle. \quad (5.29)$$

Desta forma $\mathcal{B}_{HS}$ pode ser considerado um espaço de Hilbert com produto interno de Hilbert-Schmidt, independente da escolha de base.

Sabemos que o espaço de matrizes densidade $\mathcal{S}$ da mecânica quântica - composto por matrizes hermitianas, positivas semi-definidas de traço unitário - também deverá ser visto como uma variedade (vide Sec. 3.6), com a peculiaridade de apresentar contorno suave por partes dado por matrizes densidade degeneradas, isto é, matrizes com pelo menos um autovalor zero.⁴⁹ Em analogia a Sec. 5.1, queremos um mapeamento do tipo $\rho \rightarrow \xi = \sqrt{\rho}$ para ter uma identificação clara com o que já foi apresentado nas seções anteriores.

Considere então o espaço de Hilbert $\mathcal{B}_{HS}$ equipado com o produto interno g_{ab} de Hilbert-Schmidt. Um típico vetor nesse espaço será uma matriz, que agora representar indexialmente como ξ^a pois sabemos como operar $\xi^a g_{ab} \xi^b = \text{Tr}(\xi^\dagger \xi)$, onde ξ sem índice dentro do traço é a matriz associada a ξ^a . O conjunto $\mathcal{S} = \{\rho \mid \rho = \rho^\dagger > 0 \text{ e } \text{Tr}(\rho) = 1\}$ é subconjunto de $\mathcal{B}_{HS}$, de forma que uma matriz desse subconjunto pode ser mapeada na sua raiz quadrada, $\rho \rightarrow \sqrt{\rho}$. Identificamos $\mathcal{S}$ com o mapeamento (mergulho) de ρ em sua raiz quadrada ξ de forma que $g_{ab} \xi^a \xi^b = \text{Tr}(\xi \xi) = 1$.

Uma variável estocástica em $\mathcal{S}$ é uma forma X_{ab} com valor esperado em termos do estado ξ dado por

$$E_\xi[X] = \xi^a X_{ab} \xi^b = \text{Tr}[\xi X \xi].$$

Dado o desvio da média como $\tilde{X}_{ab} = X_{ab} - g_{ab}(\xi^c X_{cd} \xi^d)$, analogamente a discussão da Sec. 5.1.1, temos que a variância de X_{ab} fica

$$\text{Var}_\xi[X] = \Delta X^2 = \xi^a \tilde{X}_{ac} \tilde{X}_b^c \xi^b,$$

e a covariância entre X_{ab} e Y_{ab} tem a forma

$$\text{Cov}_\xi[X, Y] = \xi^a \tilde{X}_{ac} \tilde{Y}_b^c \xi^b.$$

É importante ressaltar que, $X_{ac} \xi^c \xi^b Y_b^a \neq \xi^c X_{ca} \xi^b Y_b^a$ já que $\text{Tr}[X \xi \xi Y] \neq \text{Tr}[\xi X \xi Y]$. Como ξ^a agora representa uma matriz, temos que ficar muito atentos em que lado ocorrem as contrações, pois *álgebra de matrizes não é comutativa*.

Parametrizando as matrizes densidade por um conjunto de coordenadas locais $\{\theta = [\theta^i] \in \Theta \subset \mathbb{R}^n; i = 1, \dots, n\}$ temos a variedade $\mathcal{S}$ em $\mathcal{B}$ dada parametricamente por $\xi(\theta) \in C^\infty$. Analogamente à Prop. 1, escrevemos ∂_i para $\partial/\partial\theta^i$ e em coordenadas locais $[\theta^i]$ temos uma métrica Riemanniana $\mathcal{G}_{ij}$ em $\mathcal{S}$ induzida pela métrica de Hilbert-Schmidt g_{ab} de $\mathcal{B}_{HS}$,

$$\mathcal{G}_{ij} = 2\text{Tr}[\partial_i \xi \partial_j \xi] \quad (5.30)$$

A demonstração segue exatamente o mesmo raciocínio da feita na Prop. 1, posteriormente vamos discutir o fator 2 na expressão acima.

Vale destacar que no caso de $\mathcal{S}$ não precisamos inserir nenhuma estrutura complexa J_b^a , visto que o conjunto de operadores densidade já é subconjunto de $M_n(\mathbb{C})$. Além disso, como discutimos no fim da Sec. 3.6.1, tratamos $\mathcal{S}$ como variedade real.

5.3.2 Estimativa quântica no caso mistos

Suponha agora que temos o resultado de uma medida de um estimador não viesado T_{ab} , tal que $\xi^a T_{ab} \xi^b = t$, e que uma família $\xi(t)$ uniparamétrica em t caracteriza a distribuição de probabilidade dos possíveis resultados da medida. Podemos nos referir a t como o tempo que passou desde a preparação do estado inicial normalizado conhecido $\xi_0 = \xi(0)$, desejamos estimá-lo com o resultado dado.

O sistema preparado inicialmente no estado normalizado ξ_0 evolui via Hamiltoniano H_{ab} segundo a equação de movimento de Heisenberg

$$\xi_t = \exp(-iHt)\xi_0 \exp(iHt).$$

Note que derivando a equação acima no tempo obtemos

$$\dot{\xi}_t = -iH\xi_t + i\xi_t H = -i[H, \xi_t]$$

que é a dinâmica de von-Neumann para o estado ξ_t .

A condição de normalização nos dá que $\text{Tr}(\xi_t \xi_t) = 1$. Derivando essa expressão no tempo obtemos

$$\dot{\xi}^a g_{ab} \xi^b + \xi^a g_{ab} \dot{\xi}^b = 2\text{Tr}(\xi_t \dot{\xi}_t) = 0.$$

Dado que $\tilde{T}_{ab} = T_{ab} - t g_{ab}$, então diferenciando $\text{Tr}(\tilde{T} \xi_t \xi_t) = 0$ ficamos com

$$\begin{aligned} \text{Tr}(\tilde{T} \dot{\xi}_t \xi_t + \xi_t \tilde{T} \dot{\xi}_t - \xi_t \xi_t) &= 0 \\ \text{Tr}[(\tilde{T} \xi_t + \xi_t \tilde{T}) \dot{\xi}_t] &= 1. \end{aligned}$$

Evocando a desigualdade de Cauchy-Schwartz para um par de operadores hermitianos X e Y , temos

$$[\text{Tr}(XY)]^2 \geq \text{Tr}(X^2)\text{Tr}(Y^2).$$

Sendo $X = (\tilde{T} \xi_t + \xi_t \tilde{T})$ e $Y = \dot{\xi}_t$ podemos ver que

$$\begin{aligned} \text{Tr}(X^2) &= 2\text{Tr}(\tilde{T}^2 \xi_t \xi_t + \tilde{T} \xi_t \tilde{T} \xi_t) \\ &= 2\Delta T^2 + \text{Tr}(T \xi_t T \xi_t - T \xi_t t \xi_t - t \xi_t T \xi_t + t^2 \xi_t \xi_t) \\ &= 2[\Delta T^2 + \text{Tr}(T \xi_t T \xi_t - t^2)] \end{aligned}$$

definindo $\delta T^2 \equiv \text{Tr}(T \xi_t T \xi_t) - t^2$, obtemos

$$\text{Tr}(X^2) = 2(\Delta T^2 + \delta T^2).$$

Da desigualdade de Cauchy-Schwartz e dos resultados acima deduzimos a versão da desigualdade de Cramér-Rao quântica em $\mathcal{S}$

$$\Delta T^2 + \delta T^2 \geq \frac{1}{2\text{Tr}(\dot{\xi}_t \dot{\xi}_t)}$$

onde o denominador do lado direito da desigualdade é a métrica (5.30) para o caso que o estado é parametrizado por uma única coordenada local, no caso o tempo t .

A partir da dinâmica de von-Neumann nos encontramos que

$$\text{Tr}(\dot{\xi}_t \dot{\xi}_t) = 2[\text{Tr}(H^2 \xi_t \xi_t) - \text{Tr}(H \xi_t H \xi_t)].$$

Dado o mapeamento $\xi_t \rightarrow \sqrt{\rho_t}$, vemos que o lado direito de da expressão acima é duas vezes a WYSI, para um observável arbitrário H no estado quântico ρ_t , onde a WYSI é dada por

$$I_{\rho_t}(H) = \text{Tr}(H^2 \rho_t) - \text{Tr}(H \sqrt{\rho_t} H \sqrt{\rho_t}) = -\frac{1}{2} \text{Tr}([\sqrt{\rho}, H]^2). \quad (5.31)$$

Introduzindo a grandeza *informação de segundo tipo* definida como

$$\delta H^2 = \text{Tr}(H \sqrt{\rho_t} H \sqrt{\rho_t}) - [\text{Tr}(H \rho_t)]^2.$$

Reescrevemos então a equação (5.31) como

$$I_{\rho_t}(H) = (\Delta H^2 - \delta H^2).$$

Agora, a desigualdade de Cramér-Rao pode ser reescrita na forma

$$\Delta T^2 + \delta T^2 \geq \frac{1}{4(\Delta H^2 - \delta H^2)}. \quad (5.32)$$

Se $\xi_t = \xi_t^2$, segue que $\text{Tr}(H \sqrt{\rho_t} H \sqrt{\rho_t}) = [\text{Tr}(H \rho_t)]^2$, desta forma temos que $\delta H^2 = 0$, assim, $\text{Tr}(\dot{\xi}_t \dot{\xi}_t) = 4\Delta H^2$. Pelo mesmo raciocínio $\delta T^2 = 0$. Com isso, a desigualdade (5.32) se reduz ao caso puro (5.19). Entretanto note que fazer isso nos coloca na borda de $\mathcal{S}$, fora das nossas hipóteses iniciais.

Segue a justificativa por termos colocado um fator 2 na definição da métrica: a métrica $g^{(\alpha=0)}$ calculada em (3.27) corresponde a WYSI no caso uniparamétrico. Lá esta métrica possui um fator 4 multiplicativo de fato, como Brody afirma.⁶⁹ Contudo o mergulho lá é dado por $\rho \rightarrow 2\sqrt{\rho}$, que é diferente do feito aqui. Com o fator 2 conseguimos estabelecer uma compatibilidade métrica e relação (5.32), no sentido do CRB ($1/\mathcal{G}$), e recuperar forçosamente a expressão do estado puro. Note que, usando o mergulho $\rho \rightarrow 2\sqrt{\rho}$ nos cálculos, com a métrica tendo então o fator 4, quando forçamos (5.32) ao caso puro obtemos $\Delta^2 T \Delta^2 H \geq 1/8$ como em Luo.⁷⁷ Portanto dado o mergulho $\rho \rightarrow \sqrt{\rho}$, concluímos que a métrica natural, induzida pela noção de distância no espaço de Hilbert adjacente corresponderá a WYSI no espaço de estados mistos. Pela Prop. 4, essa métrica induzida é a única que apresenta α -conexão Riemanniana.

Podemos ainda colocar a desigualdade (5.32) em uma outra forma via manipulação de desigualdades:

$$\begin{aligned}\Delta T^2 \Delta H^2 + \Delta H^2 \delta T^2 - \delta T^2 \delta H^2 &\geq \Delta T^2 \Delta H^2 + \Delta H^2 \delta T^2 - \Delta T^2 \delta H^2 - \delta T^2 \delta H^2 \geq 1/4 \\ \Delta T^2 \Delta H^2 &\geq 1/4 + \delta T^2 \delta H^2 \geq 1/4 - \Delta H^2 \delta T^2 + \delta T^2 \delta H^2 \\ \Delta T^2 \Delta H^2 &\geq \frac{1}{4} + \delta T^2 \delta H^2.\end{aligned}$$

A desigualdade resultante acima é menos *tight*, porém mais simétrica.

5.3.3 Correções de ordem superior na cota da variância para o caso misto

A fim de estabelecer ordens de correção superiores as cotas da variância no caso misto, com T e H canonicamente conjugados, precisamos revisitar as cotas de Bhattacharyya generalizadas traduzindo-as para este novo contexto.

Relaxando a unitariedade do traço, seja uma família $\xi(t)$ de estados uniparamétricos em t . Dado o estimador não viesado T_{ab} para o parâmetro t definimos uma função simétrica de ξ em $\mathcal{B}_{HS}$ como

$$t(\xi) = \frac{\xi^a T_{ab} \xi^b}{\xi^c g_{cd} \xi^d} = \frac{\text{Tr}(\xi T \xi)}{\text{Tr}(\xi \xi)}.$$

Consideraremos o gradiente desta função e o restringiremos ao espaço das raízes quadradas das matrizes densidades impondo $\text{Tr}(\xi \xi) = 1$.

Lembrando que $\nabla_c = \partial/\partial \xi^c$, temos

$$\begin{aligned}\nabla_c \xi^a T_{ab} \xi^b &= \nabla_c \xi^a T_{ab} \xi^b + \xi^a T_{ab} \nabla_c \xi^b \\ &= T_{cb} \xi^b + \xi^a T_{ac} \\ &= (T_{ac} + T_{ca}) \xi^a = (T\xi + \xi T).\end{aligned}$$

Usando o mesmo raciocínio do cálculo acima, temos que o gradiente da função t em $\mathcal{B}_{HS}$ é dado por

$$\nabla_c t = \frac{(T_{ac} + T_{ca}) \xi^a - 2t g_{ac} \xi^a}{\xi^b \xi_b}.$$

Em seguida, impondo a unitariedade do traço, $\text{Tr}(\xi \xi) = 1$, ficamos com

$$\begin{aligned}\nabla_c t &= (T_{ac} + T_{ca}) \xi^a - 2t g_{ac} \xi^a \\ \nabla^c t &= (T^c_a + T_a^c) \xi^a - 2t \xi^c.\end{aligned}$$

Calculando a norma do gradiente, obtemos

$$\begin{aligned}
\|\nabla^c t\|^2 &= (T_{ac} + T_{ca})(T_b^c + T_b^c)\xi^a\xi^b \\
&\quad - 2t(T_{ac} + T_{ca})\xi^a\xi^c - 2t(T_{ab} + T_{ba})\xi^a\xi^b + 4t^2\xi^a\xi_a \\
&= T_{ca}\xi^aT_b^c\xi^b + T_{ca}\xi^a\xi^bT_b^c \\
&\quad + \xi^aT_{ac}T_b^c\xi^b + \xi^aT_{ac}\xi^bT_b^c - 4t^2 \\
&= 2\{[\text{Tr}(T^2\xi\xi) - t^2] + [\text{Tr}(T\xi T\xi) - t^2]\} \\
&= 2(\Delta T^2 + \delta T^2)
\end{aligned}$$

A dinâmica de von-Neumann pode ser reescrita como

$$\dot{\xi}^a = -iH_b^a\xi^b + i\xi^bH_b^a.$$

Dado que T é *canonicamente conjugado* a H , como pedido na Sec. 5.3.3, sabemos que $i[H, T] = 1$. Calculando a projeção de $\nabla^c t$ na direção $\dot{\xi}^a$ obtemos

$$\begin{aligned}
\nabla_a t \dot{\xi}^a &= [(T_{ac} + T_{ca})\xi^c - 2t\xi_a] [-iH_b^a\xi^b + i\xi^bH_b^a] \\
&= \text{Tr}(-iT\xi H\xi + i\xi T\xi H + iT\xi\xi H - i\xi TH\xi) \\
&= i\text{Tr}([H, T]\xi\xi) = 1.
\end{aligned}$$

Evocando Cauchy-Schwartz para a projeção acima, e com nossos resultados calculados até então, recuperamos o CRB para o caso misto (5.32). Esse raciocínio é o mesmo que foi apresentado na Sec. 5.1.3.

Proposição 17. Seja o sistema clássico ortogonal de vetores dados pelas relações (5.9) e (5.10), agora definidos no contexto de estados mistos. As cotas inferiores generalizadas de Bhattacharyya, no caso misto, de um estimador não viesado T_{ab} de uma função t pode ser expressa na forma

$$\Delta T^2 + \delta T^2 \geq \frac{1}{2} \sum_r \frac{(\nabla_a t \hat{\xi}^{(r)a})^2}{\hat{\xi}^{(r)a} \hat{\xi}_a^{(r)}} \quad (5.33)$$

Prova. Seja $\hat{T}_{ab} \equiv (T_{ab} + T_{ba}) - 2tg_{ab} = (\tilde{T}_{ab} + \tilde{T}_{ba})$. Definindo o tensor $R_{ab} \equiv \hat{T}_{ab} + \sum_r \lambda_r \xi_{(a} \hat{\xi}_{b)}$, prosseguimos com o cálculo de sua variância obtendo

$$\text{Var}_\xi[R] = \text{Var}_\xi[\hat{T}] + \sum_r \lambda_r (\xi^a \hat{T}_{ac} \hat{\xi}^{(r)c} + \hat{\xi}^{(r)c} \hat{T}_{ca} \xi^a) + \sum_r \lambda_r^2 \hat{\xi}^{(r)b} \hat{\xi}_b^{(r)}.$$

Para determinar cada valor λ_r que minimiza $\text{Var}_\xi[R]$, basta considerarmos a variância como função de variáveis $\{\lambda_r\}_{r \in \mathbb{N}^\times}$ e encontrar pontos extremos que anulam o gradiente desta função, obtendo um ponto de mínimo que possui todas entradas com o mesmo valor, a saber $\lambda_r^{\min} = -(\xi^a \hat{T}_{ac} \hat{\xi}^{(r)c} + \hat{\xi}^{(r)c} \hat{T}_{ca} \xi^a) / 2\hat{\xi}_b^{(r)} \hat{\xi}^{(r)b}$ para todo r . Quando substituimos este valor na expressão acima ficamos com

$$\min_{\{\lambda_r\}_{r \in \mathbb{N}^\times}} (\text{Var}_\xi[R]) = \text{Var}_\xi[\hat{T}] - \sum_r \frac{(\xi^a \hat{T}_{ac} \hat{\xi}^{(r)c} + \hat{\xi}^{(r)c} \hat{T}_{ca} \xi^a)^2}{4\hat{\xi}^{(r)b} \hat{\xi}_b^{(r)}}.$$

Como $\text{Var}_\xi[R] \geq 0$, chegamos em uma expressão para as cotas generalizadas para a variância de $\hat{T}_{ab}$

$$\text{Var}_\xi[\hat{T}] \geq \sum_r \frac{(\xi^a \hat{T}_{ac} \hat{\xi}^{(r)c} + \hat{\xi}^{(r)c} \hat{T}_{ca} \xi^a)^2}{4 \hat{\xi}^{(r)b} \hat{\xi}_b^{(r)}}.$$

Agora note que

$$(\xi^a \hat{T}_{ac} \hat{\xi}^{(r)c} + \hat{\xi}^{(r)c} \hat{T}_{ca} \xi^a)^2 = 4(\xi^a T_{ac} \hat{\xi}^{(r)c} + \hat{\xi}^{(r)c} T_{ca} \xi^a)^2 = 4(\nabla_a t \hat{\xi}^{(r)a})^2 \text{ e que}$$

$$\text{Var}_\xi[\hat{T}] = 2(\Delta T^2 + \delta T^2)$$

Unindo os resultados obtemos a expressão pretendida em (5.33). Naturalmente, no caso $r = 1$ recuperamos a desigualdade (5.32). $\square$

Antes de focarmos nas ordens de correção superiores vamos apresentar alguns resultados úteis que são as traduções dos enunciados da Sec. 5.3.3 para o contexto de estados mistos. Vamos omitir, a partir de então, o sub-índice t de ξ_t , pois já é claro que nossa raiz quadrada do operador densidade depende unicamente do parâmetro tempo.

Proposição 18. Seja $\xi^a(t)$ satisfazendo a dinâmica de von-Neumann, a norma de $\xi^{(n)a}$, $g_{ab} \xi^{(n)a} \xi^{(n)b}$ é independente do parâmetro t , onde $\xi^{(n)a} = d^n \xi^a / dt^n$. Segue como consequência direta que $g_{ab} \xi^{(n)a} \xi^{(n+1)b} = 0$.

Prova. Se Von-Neumann vale, então $\dot{\xi}^{(n)a} = -i H_b^a \xi^{(n)b} + i \xi^{(n)b} H_b^a$ também se aplica. Derivando a norma de $\xi^{(n)a}$ no tempo temos

$$\begin{aligned} d/dt[g_{ab} \xi^{(n)a} \xi^{(n)b}] &= 2g_{ab} \xi^{(n)a} \dot{\xi}^{(n)b} = 2g_{ab} \xi^{(n)a} \xi^{(n+1)b} \\ &= \xi_b^{(n)} (-i H_c^a \xi^{(n)c} + i \xi^{(n)c} H_c^a) = \text{Tr}(-i \xi^{(n)} H \xi^{(n)} + i \xi^{(n)} H \xi^{(n)}) = 0, \end{aligned}$$

completando a prova. $\square$

Um detalhe importante é que podemos escrever a equação de von-Neumann como $\dot{\xi} = -i[\tilde{H}, \xi]$, onde $\tilde{H}_{ab} = H_{ab} - \text{Tr}(H \xi \xi)$. Dito isso é fácil generalizar a seguinte relação

$$\xi^{(n)} = \text{mod}_{(-i)}[n] \cdot \text{Ad}_{(\tilde{H})}^n[\xi], \quad (5.34)$$

onde

$$\begin{aligned} \text{mod}_{(-i)}[n] &\implies n = \bar{1} \rightarrow -i \\ &\quad n = \bar{2} \rightarrow -1 \\ &\quad n = \bar{3} \rightarrow i \\ &\quad n = \bar{4} \rightarrow +1 \end{aligned}$$

e

$$\text{Ad}_{(\tilde{H})}^n[\cdot] = [\dots [\tilde{H}, [\tilde{H}, [\tilde{H}, [\tilde{H}, \cdot]]]] \dots].$$

Proposição 19. Seja T_{ab} canonicamente conjugado a H_{ab} e estimador não viesado para o parâmetro t . Então

$$T_{ab}\xi^{(n)a}\xi^{(n)b} = tg_{ab}\xi^{(n)a}\xi^{(n)b} + \kappa, \quad (5.35)$$

onde κ é uma constante. Assim, para todo n , $\tilde{T}_{ab}\xi^{(n)a}\xi^{(n)b} = \kappa$ é uma constante de movimento ao longo da trajetória $\xi(t)$ respeitando a dinâmica de von-Neumann.

Prova. Se T é canonicamente conjugado a H então $ig_{ab} = (T_{ac}H^c_b - H_a{}^cT_{cb})$. Segue que

$$g_{ab}\xi^{(n)a}\xi^{(n)b} = (-i\xi^{(n)a}T_{ac}H^c_b\xi^{(n)b} + i\xi^{(n)a}H_a{}^cT_{cb}\xi^{(n)b}).$$

Derivando o valor esperado de T no estado $\xi^{(n)}$ ficamos com

$$\begin{aligned} d/dt[T_{ab}\xi^{(n)a}\xi^{(n)b}] &= T_{ab}(\dot{\xi}^{(n)a}\xi^{(n)b} + \xi^{(n)a}\dot{\xi}^{(n)b}) \\ &= (-i\xi^{(n)a}T_{ac}H^c_b\dot{\xi}^{(n)b} + i\dot{\xi}^{(n)a}H_a{}^cT_{cb}\xi^{(n)b}) = g_{ab}\xi^{(n)a}\xi^{(n)b}. \end{aligned}$$

Integrando a expressão acima obtemos o resultado requerido

$$\begin{aligned} T_{ab}\xi^{(n)a}\xi^{(n)b} &= tg_{ab}\xi^{(n)a}\xi^{(n)b} + \kappa \\ \tilde{T}_{ab}\xi^{(n)a}\xi^{(n)b} &= \kappa, \end{aligned}$$

onde κ é uma constante independente de t e $\tilde{T}_{ab}\xi^{(n)a}\xi^{(n)b}$ é uma constante de movimento. $\square$

Proposição 20. Seja T_{ab} canonicamente conjugado a H_{ab} e estimador não viesado para o parâmetro t . Então para inteiros ímpares n , com $m = (n - 1)/2$, temos

$$(T_{ab} + T_{ba})\xi^{(n)a}\xi^b = (-1)^m n g_{ab}\xi^{(m)a}\xi^{(m)b}. \quad (5.36)$$

Prova. Como vale o resultado da Prop. 19, o esquema da prova é idêntico ao apresentado no do Lema. 14. $\square$

Em posse desses resultados estamos prontos para deduzir correções de ordem superior no caso misto para relações de incerteza do tipo Heisenberg dados observáveis canonicamente conjugados. Lembrando que estamos no contexto de estimativa do parâmetro t no caso de uma família de raízes quadradas de matrizes densidade uniparamétricas $\xi^a(t)$ que obedecem a dinâmica de von-Neumann, dado Hamiltoniano H_{ab} . O observável T_{ab} é o estimador não viesado do parâmetro t tomado canonicamente conjugado ao Hamiltoniano.

Vamos então considerar correções até a terceira ordem para a cota da variância de T_{ab} que emergem quando expandimos $\nabla_a t$ no sistema clássico ortogonal de vetores, dados por $\dot{\xi}^a$, $\hat{\xi}^{(2)a}$ e $\hat{\xi}^{(3)a}$. Por (5.33), a cota da variância fica

$$\Delta T^2 + \delta T^2 \geq \frac{(\dot{\xi}^a \nabla_a t)^2}{2\dot{\xi}^b \dot{\xi}_b} + \frac{(\hat{\xi}^{(2)a} \nabla_a t)^2}{2\hat{\xi}^{(2)b} \hat{\xi}_b^{(2)}} + \frac{(\hat{\xi}^{(3)a} \nabla_a t)^2}{2\hat{\xi}^{(3)b} \hat{\xi}_b^{(3)}}, \quad (5.37)$$

onde $\hat{\xi}^{(2)a}$ e $\hat{\xi}^{(3)a}$ são dados por

$$\hat{\xi}^{(2)a} = \ddot{\xi}^a - \frac{(\ddot{\xi}^b \dot{\xi}_b)}{(\dot{\xi}^c \dot{\xi}_c)} \dot{\xi}^a - (\ddot{\xi}^b \xi_b) \xi^a$$

e

$$\hat{\xi}^{(3)a} = \ddot{\xi}^a - \frac{(\ddot{\xi}^b \ddot{\xi}_b)}{(\ddot{\xi}^c \ddot{\xi}_c)} \ddot{\xi}^a - \frac{(\ddot{\xi}^b \dot{\xi}_b)}{(\dot{\xi}^c \dot{\xi}_c)} \dot{\xi}^a - (\ddot{\xi}^b \xi_b) \xi^a.$$

O termo correspondente a $r = 1$ em (5.37) já foi calculado na derivação de (5.32). Vamos proceder então com o termo de segunda ordem. Os resultados (5.34) e Prop. 18 nos dão que $\ddot{\xi} = -[\tilde{H}, [\tilde{H}, \xi]]$, $\ddot{\xi}^a \dot{\xi}_a = 0$, $\ddot{\xi}^a \xi_a = -2I_\rho(\tilde{H})$. Portanto, $\hat{\xi}^{(2)a} = -[\tilde{H}, [\tilde{H}, \xi]] + 2I_\rho(\tilde{H})\xi$. Assim sendo o numerador do termo de segunda ordem fica $(\hat{\xi}^{(2)a} \nabla_a t)^2 = \left\{ \text{Tr}[(T\xi - \xi T)(-[\tilde{H}, [\tilde{H}, \xi]] + 2I_\rho(\tilde{H})\xi)] \right\}^2$ que vai depender explicitamente da escolha do estimador T . Em conta dessa dependência, como já salientado nas seções referentes ao caso puro, vamos descartar esse termo, pois queremos apenas termos de correção que não dependam da escolha do estimador para contribuir na generalização da cota.

Agora, passando para o termo que envolve $\hat{\xi}^{(3)a}$, segundo a Prop. 18, $\ddot{\xi}^a \ddot{\xi}_a = \ddot{\xi}^a \dot{\xi}_a = 0$ segue que $\ddot{\xi}^a \xi_a = -\dot{\xi}^a \dot{\xi}_a$, então $\ddot{\xi}^a \xi_a = 0$. Pela (5.34) $\ddot{\xi}^a = i[\tilde{H}, [\tilde{H}, [\tilde{H}, \xi]]] = i([\tilde{H}^3, \xi] + 3[\tilde{H}\xi\tilde{H}, \tilde{H}])$. Após alguns cálculos encontramos que $\ddot{\xi}^a \dot{\xi}_a = -2\text{Tr}(\tilde{H}^4 \xi \xi - 4\tilde{H}^3 \xi \tilde{H} \xi + 3\tilde{H}^2 \xi \tilde{H}^2 \xi)$. Logo

$$\hat{\xi}^{(3)} = i \left\{ ([\tilde{H}^3, \xi] + 3[\tilde{H}\xi\tilde{H}, \tilde{H}]) - \frac{\langle \tilde{H}^4 \rangle_{mix}}{\langle \tilde{H}^2 \rangle_{mix}} \right\}$$

e

$$\hat{\xi}^{(3)a} \hat{\xi}_a^{(3)} = \langle \tilde{H}^6 \rangle_{mix} - \frac{\langle \tilde{H}^4 \rangle_{mix}^2}{\langle \tilde{H}^2 \rangle_{mix}}, \quad (5.38)$$

onde agora definimos

$$\begin{aligned} \langle \tilde{H}^2 \rangle_{mix} &:= 2I_\rho(\tilde{H}) = 2\text{Tr}(\tilde{H}^2 \xi \xi - \tilde{H} \xi \tilde{H} \xi) = 2(\Delta H^2 + \delta H^2) = \text{Tr}(\dot{\xi} \dot{\xi}) \\ \langle \tilde{H}^4 \rangle_{mix} &:= 2\text{Tr}(\tilde{H}^4 \xi \xi - 4\tilde{H}^3 \xi \tilde{H} \xi + 3\tilde{H}^2 \xi \tilde{H}^2 \xi) = \text{Tr}(\ddot{\xi} \ddot{\xi}) \\ \langle \tilde{H}^6 \rangle_{mix} &:= \text{Tr}(\tilde{H}^6 \xi \xi - 6\tilde{H}^5 \xi \tilde{H} \xi + 15\tilde{H}^4 \xi \tilde{H}^2 \xi - 10\tilde{H}^3 \xi \tilde{H}^3 \xi) = \text{Tr}(\xi^{(3)} \xi^{(3)}). \end{aligned}$$

Para o numerador do termo de terceira ordem

$$\hat{\xi}^{(3)a} \nabla_a t = i \text{Tr} \left\{ [T, \tilde{H}^3] \xi \xi + 3(\xi [\tilde{H}^2, T] \xi \tilde{H} + \xi [T, \tilde{H}] \xi \tilde{H}^2) - \xi [T, \tilde{H}] \xi \frac{\langle \tilde{H}^4 \rangle_{mix}}{\langle \tilde{H}^2 \rangle_{mix}} \right\},$$

vemos que ele envolve comutadores entre T e H^k , $k \in \mathbb{N}$. É fácil demonstrar por indução finita que $[\tilde{H}^k, T] = -ki\tilde{H}^{k-1}$, $k \in \mathbb{N}$. Dado esse fato temos que

$$\hat{\xi}^{(3)a} \nabla_a t = \frac{\langle \tilde{H}^4 \rangle_{mix}^2}{\langle \tilde{H}^2 \rangle_{mix}} - 3 \langle \tilde{H}^2 \rangle_{mix}. \quad (5.39)$$

Parece ser regra que termos pares dependam explicitamente da escolha de um estimador T e que os termos ímpares envolvam comutadores entre T e H^k , $k \in \mathbb{N}$. Já termos ímpares,

em geral, não dependem da escolha de T . Entretanto estas afirmações são pretensiosas e carecem de demonstração.

Unindo os resultados (5.32), (5.38) e (5.39), obtemos a seguinte correção para o CRB no caso misto que só depende da dinâmica de $\xi(t)$ gerada por H :

Proposição 21. Se T e H são canonicamente conjugados, então a seguinte cota

$$(\Delta T^2 + \delta T^2)(\Delta H^2 - \delta H^2) \geq \frac{1}{4} \left[1 + \frac{(\langle \tilde{H}^4 \rangle_{mix} - 3 \langle \tilde{H}^2 \rangle_{mix}^2)^2}{\langle \tilde{H}^6 \rangle_{mix} \langle \tilde{H}^2 \rangle_{mix} - \langle \tilde{H}^4 \rangle_{mix}^2} \right] \quad (5.40)$$

se aplica ao produto acima, com estado $\rho(t)$ seguindo dinâmica von-Neumann.

Se formosmos $\xi = \xi^2$, $\langle \tilde{H}^4 \rangle_{mix} \neq \langle \tilde{H}^4 \rangle$ por exemplo. Isso só será verdade com $\langle \tilde{H}^2 \rangle_{mix} = \langle \tilde{H}^2 \rangle$. Então, (5.40) não volta para (5.27) apenas impondo $\xi = \xi^2$, ao contrário do que foi dito por Brody.⁶⁹

Espaço de estados quântico é um conjunto convexo com bordo contínuo por partes. Ao longo do texto trabalhamos apenas com o miolo $\mathcal{S}$, variedade real, sobre hipóteses específicas que foram adotadas ao longo de todo texto. Estados puros vivem na borda deste espaço e não podem ser representados por nenhuma combinação convexa de outros pontos extremais, além do fato deste espaço possuir estrutura complexa e ser identificado com o espaço projetivo complexo. Seria muito otimista pensar que a passagem de uma abordagem feita em $\mathcal{S}$ se reduziria ao caso puro apenas fazendo $\rho = \rho^2$, lembrando também que na Sec. 4.1 comentamos que para aplicar o ao caso puro Teorema de MCP deveríamos passar por um processo chamado extensão radial, reduzindo todas as métricas à Fubini-Study. Acreditamos que a passagem para o caso puro deva usar uma outra abordagem conhecida como a inversa de Moore-Penrose.⁷⁸ Neste ponto só podemos apenas dizer que se os estados forem puros a abordagem feita na Sec. 5.2 é que deve ser usada, caso contrário, se forem mistos a abordagem empregada nesta seção deve ser usada.

6 CONCLUSÃO

Nesta dissertação apresentamos estudos desenvolvidos acerca de variedades estatísticas clássica e quântica, onde o caso clássico apresenta estrutura dualística $(g, \nabla^\alpha, \nabla^{-\alpha})$ correspondente a métrica de Fisher e $\pm\alpha$ -conexões vinda de α -divergências globalmente definidas através de funções convexas, que quando estendida ao caso quântico produz uma família de métricas $g^{(\alpha)}$ para direções não-comutativas do espaço de estados quântico $\mathcal{S}$. Tais métricas $g^{(\alpha)}$ se relacionam com as métricas contrativas monotônicas permitidas pelo Teorema de MCP para espaços quânticos via mapeamento operador função convexa em operador função monotônica, produzindo assim métricas Riemannianas contrativas sob ação de mapa CPTP e invariantes sobre ação do grupo unitário.

Ao aplicarmos métodos geométricos a um problema de inferência estatística, que consiste em calcular correções de ordem superior para o CRB associado a um estimador uniparamétrico utilizando da estrutura de produto interno do espaço de Hilbert adjacente, pudemos ver várias ligações com os tópicos anteriores. O mergulho escolhido foi aquele que leva elementos uniparamétricos do espaço de fundo em suas raízes quadrada, pois nesse caso ficava fácil impor a condição de normalização com o produto interno do espaço de Hilbert. Quando nos restringimos a variedade estatística (impondo normalização) recuperamos as métricas correspondentes a Fisher no caso clássico, Fubini-Study no caso quântico puro e $g^{\alpha=0}$ no caso quântico misto. Pudemos ver também que esta métrica induzida pelo produto interno, dado o mergulho do tipo $x \rightarrow \sqrt{x}$, tem conexão Riemanniana única igual a 0-conexão. O CRB para um estimador do parâmetro aparece do fato que um vetor pertencente ao espaço de Hilbert de fundo sempre terá norma maior que a sua componente ortogonalmente projetada na variedade estatística, que também é o raciocínio por trás das cotas de Bhattacharyya.

Vale destacar que a métrica correspondente a $\alpha = 0$ se relaciona com a função $f_{WY}(u)$ de Wigner-Yanase prevista pelo Teorema MCP, que no caso uniparamétrico ambas são iguais a informação de Wigner-Yanase a menos de um fator constante. Neste ponto advogamos pelo uso da informação de Wigner-Yanase como métrica natural em problemas de inferência estatística,^{79,80} afinal, do ponto de vista matemático é a única métrica com α -conexão Riemanniana e apresenta mergulho simples de se trabalhar. Fisicamente o CRB correspondente a ela leva em conta as coerências quânticas e se reduz ao caso puro apenas forçando $\rho = \rho^2$. Poderíamos também querer trabalhar com métricas de Wigner-Yanase-Dyson-Lieb correspondentes a $\alpha \in [-1, 1]$, que são garantidamente contrativas por Hasegawa-Petz-Lieb, mas lembre-se que elas podem ter mergulhos mais intrincados e não possuem α -conexão Riemanniana análoga a estrutura dualística clássica. Pode decorrer daqui algum trabalho experimental que teste a viabilidade da relação de incerteza que

emana da escolha da Wigner-Yanase como métrica do espaço de estados.

Quanto as correções, conseguimos obter correções de ordem superior legítimas, positivas e independentes da escolha de um estimador, para cotas do tipo Heisenberg no caso de estados mistos. Vimos que as correções do caso quântico de estado puro são parecidas na forma com as do caso misto. Porém as correções no caso misto não se reduzem ao caso puro apenas forçando $\rho = \rho^2$. Como já apontamos no capítulo 4, passagem do caso misto para o puro pode ser perigosa e não trivial, de forma que não basta apenas impor $\rho = \rho^2$. Lembrando, espaço de estados quânticos é um conjunto convexo, onde ao longo do texto trabalhamos apenas com o miolo $\mathcal{S}$ com hipóteses de dimensão finita e espectro não degenerado, já estados puros vivem na borda contínua por partes deste espaço e não podem ser representados por nenhuma combinação convexa de outros pontos extremais. Acreditamos que a passagem para o caso puro deva usar uma outra abordagem conhecida como a inversa de Moore-Penrose,^{78,81} sendo uma possível continuidade deste trabalho.

Podemos também atentar para um outro fato, que podemos deixar como sugestão para trabalhos futuros. O que aconteceria se não pedirmos a positividade das métricas? Sabemos que geometrias importantes na Física tem suas métricas não satisfazendo positividade. A questão é, a não positividade de métricas em mecânica quântica traria a tona alguma nova descrição física relevante? Afinal, essa escolha de positividade é arbitrária.

REFERÊNCIAS

- 1 SZEKERES, P. **A course in modern mathematical physics: groups, Hilbert spaces, and differential geometry**. 4th. ed. New York: Cambridge University Press, 2012. ISBN 9780521829601.
- 2 AMARI, S.; NAGAOKA, H. **Methods of information geometry**. Oxford: Oxford University Press, 2000. ISBN 9780821843024.
- 3 BRODY, D.; HUGHSTON, L. Geometry of quantum statistical inference. **Physical Review Letters**, American Physical Society, v. 77, n. 14, p. 2851–2854, 1996.
- 4 DEDEKIND, R. **Bernhard Riemann's Lebenslauf**, 2007. Disponível em: <http://archiv.ub.uni-heidelberg.de/volltextserver/13006/1/riemann.pdf>. Acesso em 30 abr. 2018.
- 5 KLEIN, F. A comparative review of recent researches in geometry. **Bulletin of the American Mathematical Society**, v. 2, n. 10, p. 215–249, 1893.
- 6 STILLWELL, J. **Sources of hyperbolic geometry**. Providence: American Mathematical Society, 1996. ISBN 9780821809228.
- 7 HEATH, T. L. et al. **The thirteen books of Euclid's elements**. USA: Courier Corporation, 1956. ISBN 9780486600888.
- 8 RINDLER, W. **Introduction to special relativity**. 2nd. ed. Oxford: Oxford University Press, 1991. ISBN 9780198539520.
- 9 KOPCZYNSKI, W.; TRAUTMAN, A. **Spacetime and gravitation**. Chichester: John Wiley, 1992. ISBN 9780471921868.
- 10 SACHS, R. K.; WU, H. **General relativity for mathematicians**. New York: Springer Verlag, 1977. ISBN 9781461299059.
- 11 WALD, R. M. **General relativity**. Chicago: University of Chicago Press, 2010. ISBN 9780226870373.
- 12 GOLDSTEIN, H.; POOLE, C.; SAFKO, J. **Classical mechanics**. 2nd. ed. Reading, Mass: Addison-Wesley, 1980. ISBN 9780201029185.
- 13 RUDIN, W. **Funcional analysis**. 2nd. ed. New York: Mcgraw Hill, 1991. ISBN 9780070619883.
- 14 NIELSEN, M.; I., C. **Quantum computation and quantum information**. 10th. ed. Cambridge: Cambridge University Press, 2011. ISBN 9781107002173.
- 15 ARUTE, F. et al. Quantum supremacy using a programmable superconducting processor. **Nature**, Nature Publishing Group, v. 574, n. 7779, p. 505–510, 2019.
- 16 LANGE, G. D. et al. Realization of microwave quantum circuits using hybrid superconducting-semiconducting nanowire josephson elements. **Physical Review Letters**, APS, v. 115, n. 12, p. 127002, 2015.

- 17 LANYON, B. et al. Measurement-based quantum computation with trapped ions. **Physical Review Letters**, APS, v. 111, n. 21, p. 210501, 2013.
- 18 SCHINDLER, P. et al. A quantum information processor with trapped ions. **New Journal of Physics**, IOP Publishing, v. 15, n. 12, p. 123012, 2013.
- 19 SLODIČKA, L. et al. Atom-atom entanglement by single-photon detection. **Physical Review Letters**, APS, v. 110, n. 8, p. 083603, 2013.
- 20 STUTE, A. et al. Quantum-state transfer from an ion to a photon. **Nature Photonics**, Nature Publishing Group, v. 7, n. 3, p. 219, 2013.
- 21 MONROE, C. et al. Demonstration of a fundamental quantum logic gate. **Physical Review Letters**, APS, v. 75, n. 25, p. 4714, 1995.
- 22 BRUNE, M. et al. Observing the progressive decoherence of the “meter” in a quantum measurement. **Physical Review Letters**, APS, v. 77, n. 24, p. 4887, 1996.
- 23 PIRES, D. P. **Geometria da informação quântica**: uma abordagem geral acerca do tempo de evolução. 2017. 221 p. Tese (Doutorado) — Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2017.
- 24 FUCHS, C. A. **Distinguishability and accessible information in quantum theory**. 1996. Disponível em: <<https://arxiv.org/pdf/quant-ph/9601020.pdf>>. Acesso em: 6 nov. 2019.
- 25 WATROUS, J. **The theory of quantum information**. Cambridge: Cambridge University Press, 2018. ISBN 9781107180567.
- 26 WOOTTERS, W. K. Statistical distance and hilbert space. **Physical Review D**, APS, v. 23, n. 2, p. 357, 1981.
- 27 MOROZOVA, E.; CHENTSOV, N. Markov invariant geometry on state manifolds. **Itogi Nauki i Tekhniki. Seriya "Sovremennye Problemy Matematiki. Noveishie Dostizheniya"**, v. 36, p. 69–102, 1989. (in Russian).
- 28 BRAUNSTEIN, S. L.; CAVES, C. M. Statistical distance and the geometry of quantum states. **Physical Review Letters**, APS, v. 72, n. 22, p. 3439, 1994.
- 29 PETZ, D. Monotone metrics on matrix spaces. **Linear Algebra and its Applications**, Elsevier, v. 244, p. 81–96, 1996. doi: 10.1016/0024-3795(94)00211-8.
- 30 LEE, J. **Riemannian manifolds**: an introduction to curvature. Berlin: Springer, 1997. ISBN 9780387983226.
- 31 FELLER, W. **An introduction to probability theory and its applications**. 3rd. ed. New York: John Wiley, 1968. ISBN 9780471257080.
- 32 KOLMOGOROV, A. **Foundations of the theory of probability**. Eastford: Martino Fine Books, 2013. ISBN 9781614275145.
- 33 DOOB, J. **Stochastic process**. New York: John Wiley, 1990. ISBN 9780471523697.
- 34 GUT, A. **Probability**: a graduate course. Berlin: Springer, 2005. ISBN 9780387228334.

- 35 JENČOVÁ, A. Geodesic distances on density matrices. **Journal of Mathematical Physics**, AIP, v. 45, n. 5, p. 1787–1794, 2004.
- 36 EGUCHI, S. Second order efficiency of minimum contrast estimation in a curved exponential family. **Annals of Statistics**, JSTOR, v. 11, n. 3, p. 793–803, 1983.
- 37 _____. A characterization of second order efficiency in a curved exponential family. **Annals of the Institute of Statistical Mathematics**, Springer, v. 36, n. 2, p. 199–206, 1984.
- 38 _____. Geometry of minimum contrast. **Hiroshima Mathematical Journal**, v. 22, n. 3, p. 631–647, 1992.
- 39 MATUMOTO, T. Any statistical manifold has a contrast function-on the c^3 -functions taking the minimum at the diagonal of the product manifold. **Hiroshima Mathematical Journal**, v. 23, n. 2, p. 327–332, 1993.
- 40 CSISZÁR, I. Information-type measures of difference of probability distributions and indirect observation. **Studia Scientiarum Mathematicarum Hungarica**, v. 2, p. 229–318, 1967.
- 41 _____. On topological properties of f-divergence. **Studia Scientiarum Mathematicarum Hungarica**, v. 2, p. 329–339, 1967.
- 42 VAJDA, I. **Theory of statistical inference and information**. Dordrecht: Kluwer Academic Pub, 1989.
- 43 CHENTSOV, N. N. **Statistical decision rules and optimal inference**. Providence: American Mathematical Society, 1982. (Translations of mathematical monographs, v. 53).
- 44 RÉNYI, A. On measures of entropy and information. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 4., 1961, Berkeley. **Proceedings...** Berkeley: University of California Press, 1961. v. 1, p. 547–561.
- 45 CHERNOFF, H. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. **Annals of Mathematical Statistics**, v. 23, n. 4, p. 493–507, 1952.
- 46 TSALLIS, C. Possible generalization of boltzmann-gibbs statistics. **Journal of Statistical Physics**, Springer, v. 52, n. 1-2, p. 479–487, 1988.
- 47 CURADO, E.; TSALLIS, C. Generalized statistical mechanics: connection with thermodynamics. **Journal of Physics A: mathematical and general**, v. 24, n. 2, p. L69, 1991.
- 48 TSALLIS, R. M. C.; PLASTINO, A. The role of constraints within generalized nonextensive statistics. **Physica A: statistical mechanics and its applications**, Elsevier, v. 261, n. 3-4, p. 534–554, 1998.
- 49 BENGTSSON, I.; ZYCZKOWSKI, K. **Geometry of quantum states: an introduction to quantum entanglement**. Cambridge: Cambridge University Press, 2017. ISBN 9780521891400.

- 50 HUYBRECHTS, D. **Complex geometry**: an introduction. Berlin: Springer Science & Business Media, 2006. ISBN 9783540212904.
- 51 LEE, J. **Introduction to smooth manifolds**. Berlin: Springer, 2002. ISBN 9781441999818.
- 52 KOLAŘ P. W. MICHOR, J. S. I. **Natural operations in differential geometry**. Berlin: Springer, 1993. ISBN 9783662029503.
- 53 HASEGAWA, H. α -divergence of the non-commutative information geometry. **Reports on Mathematical Physics**, Elsevier, v. 33, n. 1-2, p. 87–93, 1993.
- 54 WILDE, M. **Quantum information theory**. Cambridge: Cambridge University Press, 2013. ISBN 9781107034259.
- 55 AMARAL A. BARAVIEIRA, M. T. C. B. Mecânica quântica para matemáticos em formação. In: COLÓQUIO BRASILEIRO DE MATEMÁTICA, 28., 2011. Rio de Janeiro. **Resumos...** Rio de Janeiro: IMPA, 2011. p. 210.
- 56 PETZ, D. Quasi-entropies for finite quantum systems. **Reports on Mathematical Physics**, Elsevier, v. 23, n. 1, p. 57–65, 1986.
- 57 BENDAT, J.; SHERMAN, S. Monotone and convex operator functions. **Transactions of the American Mathematical Society**, JSTOR, v. 79, n. 1, p. 58–71, 1955.
- 58 LIEB, E. **Convex trace functions and the Wigner-Yanase-Dyson conjecture**. 1973. Disponível em: <http://www.numdam.org/article/RCP25_1973__19__A4_0.pdf>. Acesso em 23 jan. 2019.
- 59 NIKOUFAR, I.; EBADIAN, A.; GORDJI, M. E. The simplest proof of lieb concavity theorem. **Advances in Mathematics**, Elsevier, v. 248, p. 531–533, 2013.
- 60 WIGNER, E.; YANASE, M. Information contents of distributions. In: WIGHTMAN, A. S. (Ed.) **Part I: particles and fields. part II: foundations of quantum mechanics**. Berlin: Springer, 1997. p. 452–460.
- 61 JENČOVÀ, A. Generalized relative entropies as contrast functionals on density matrices. **International Journal of Theoretical Physics**, Springer, v. 43, n. 7, p. 1635–1649, 2004.
- 62 HANSEN, F. Characterization of symmetric monotone metrics on the state space of quantum systems. **Quantum Information and Computation**, v. 6, n. 6, p. 597, 2006.
- 63 _____. Metric adjusted skew information. **Proceeding of the National Academy of Science**, National Acad Sciences, v. 105, n. 29, p. 9909–9916, 2008.
- 64 PETZ, D.; SUDÀR, C. Geometries of quantum states. **Journal of Mathematical Physics**, AIP, v. 37, n. 6, p. 2662–2673, 1996.
- 65 HIAI, F.; MOSONYI, M. Different quantum f-divergences and the reversibility of quantum operations. **Reviews in Mathematical Physics**, World Scientific, v. 29, n. 07, p. 1750023., 2017.

- 66 LESNIEWSKI, A.; RUSKAI, M. Monotone riemannian metrics and relative entropy on noncommutative probability spaces. **Journal of Mathematical Physics**, AIP, v. 40, n. 11, p. 5702–5724, 1999.
- 67 KUBO, M. T. R.; HASHITSUME, N. **Statistical Physics II** -nonequilibrium statistical mechanics. New York: Springer-Verlag, 1991. ISBN 9783540538332.
- 68 BRODY, D.; HUGHSTON, L. Statistical geometry in quantum mechanics. **Philosophical Transactions of the Royal Society A**, 1996. Disponível em: <<https://arxiv.org/pdf/gr-qc/9701051.pdf>>. Acesso em: 30 abr. 2018.
- 69 BRODY, D. Information geometry of density matrices and state estimation. **Journal of Physics A: mathematical and theoretical**, IOP Publishing, v. 44, n. 25, p. 252002, 2011.
- 70 BRODY, D. C. A note on exponential families of distributions. **Journal of Physics A: mathematical and theoretical**, IOP Publishing, v. 40, n. 30, p. F691, 2007.
- 71 BHATTACHARYYA, A. On some analogues of the amount of information and their use in statistical estimation. **Sankhyā: the Indian journal of statistics**, JSTOR, v. 8, n. 1, p. 1–14, 1946.
- 72 _____. On some analogues of the amount of information and their use in statistical estimation (contd.). **Sankhyā: the Indian journal of statistics**, JSTOR, v. 8, n. 3, p. 201–218, 1947.
- 73 _____. On some analogues of the amount of information and their use in statistical estimation (concluded). **Sankhyā: the Indian journal of statistics**, JSTOR, v. 8, n. 4, p. 315–328, 1948.
- 74 SHIFRIN, T. **Differential geometry: a first course in curves and surfaces**. 2015. University of Georgia, 2015. Disponível em: <<http://alpha.math.uga.edu/~shifrin/ShifrinDiffGeo.pdf>>. Acesso em: 30 abr. 2018.
- 75 STUART, A.; ORD, J. **Kendall's advanced theory of statistics**. 6th. ed. New York: Wiley, 2010. ISBN 9780470665305.
- 76 HAEGEMAN, J. et al. Geometry of matrix product states: metric, parallel transport, and curvature. **Journal of Mathematical Physics**, AIP, v. 55, n. 2, p. 021902, 2014.
- 77 LUO, S. Heisenberg uncertainty relation for mixed states. **Physical Review A**, APS, v. 72, n. 4, p. 042110, 2005.
- 78 PENROSE, R. A generalized inverse for matrices. **Mathematical Proceedings of the Cambridge Philosophical Society**, v. 51, n. 3, p. 406–413, 1955.
- 79 LI, J.; FEI, S.-M. Uncertainty relation based on Wigner–Yanase–Dyson skew information with quantum memory. **Entropy**, Multidisciplinary Digital Publishing Institute, v. 20, n. 2, p. 132, 2018.
- 80 YANAGI, K. Wigner-Yanase-Dyson skew information and uncertainty relation. **Journal of Physics: conference series**, v. 201, n. 1, p. 012015, 2010.

- 81 BARATA, J. C. A.; HUSSEIN, M. S. The moore–penrose pseudoinverse: a tutorial review of the theory. **Brazilian Journal of Physics**, Springer, v. 42, n. 1-2, p. 146–165, 2012.
- 82 LEMOS, N. **Convite à física matemática**. São Paulo: Livraria da Física, 2013. ISBN 9788578611927.
- 83 WILLARD, S. **General topology**. Mineola: Dover Publications, 2004. ISBN 9780486434797.
- 84 JEEVANJEE, N. **An introduction to tensors and group theory for physicists**. Berlin: Springer, 2011. ISBN 9780817647148.
- 85 LANG, S. **Algebra**. New York: Springer-Verlag, 2002. ISBN 9780387953854.
- 86 ARTIN, M. **Algebra**. Upper Saddle River: Prentice Hall, 1991. ISBN 9780130047632.
- 87 ZHENG, S. **Fisher information and Cramér-Rao bound**. Disponível em: http://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/Fisher_info.pdf. Acesso em: 25 set. 2019.
- 88 BASTOS, W. P. **Estudo da atingibilidade do limite de Cramér-Rao em metrologia quântica: surgimento do limite de heisenberg**. 2016. 120 p. Tese (Doutorado) — Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.
- 89 ESCHER, B. d. M. **Estimação de parâmetros em sistemas quânticos abertos**. 2011. 109 p. Tese (Doutorado) — Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.
- 90 TADDEI, M. M. **Quantum speed limits for general physical processes**. 2014. 115 p. Tese (Doutorado) — Instituto de Física, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2014.
- 91 GIOVANNETTI, V.; LLOYD, S.; MACCONE, L. Quantum metrology. **Physical Review Letters**, APS, v. 96, n. 1, p. 010401, 2006.
- 92 _____. Quantum-enhanced measurements: beating the standard quantum limit. **Science**, American Association for the Advancement of Science, v. 306, n. 5700, p. 1330–1336, 2004.
- 93 _____. Advances in quantum metrology. **Nature Photonics**, Nature Publishing Group, v. 5, n. 4, p. 222, 2011.
- 94 HELSTROM, C. Quantum detection and estimation theory. **Journal of Statistical Physics**, Springer, v. 1, n. 2, p. 231–252, 1969.
- 95 _____. Minimum mean-squared error of estimates in quantum statistics. **Physics Letters A**, Elsevier, v. 25, n. 2, p. 101–102, 1967.
- 96 _____. Estimation of a displacement parameter of a quantum system. **International Journal of Theoretical Physics**, Springer, v. 11, n. 6, p. 357–378, 1974.
- 97 _____. **Quantum detection and estimation theory**. Cambridge: Cambridge Academic Press, 1976. ISBN 9780080956329.

-
- 98 SHITARA, T.; UEDA, M. Determining the continuous family of quantum fisher information from linear-response theory. **Physical Review A**, APS, v. 94, n. 6, p. 062316, 2016.

APÊNDICE A – ELEMENTOS DE TOPOLOGIA GERAL E TENSORES

Neste apêndice vamos apresentar conceitos básicos de topologia geral e tensores, a fim de fornecer alguns pré-requisitos necessários para o leitor acompanhar a introdução matemática do Cap. 2.

Utilizamos como base os textos de Lemos⁸² e Szekeres.¹ Para o leitor interessado, recomendamos como leitura especializada de topologia o texto de Willard.⁸³ No caso de tensores, uma abordagem familiar para físicos pode ser vista no texto de Jeevanjee,⁸⁴ que trás formalidade juntamente com aplicações.

A.1 Topologia

A.1.1 Espaços topológicos

Dado um conjunto X não vazio, uma *topologia* em X consiste de uma coleção de subconjuntos τ de X , chamados *conjuntos abertos* (ou simplesmente *abertos*), que satisfazem

- i) O conjunto todo X e o vazio $\emptyset$ são membros de τ .
- ii) Se $\mathcal{O} \in \tau$ e $\mathcal{V} \in \tau$ então $\mathcal{O} \cap \mathcal{V} \in \tau$.
- iii) Se I é conjunto arbitrário de índices e $\mathcal{O}_i$ é uma família de abertos para cada $i \in I$, então $\bigcup_{i \in I} \mathcal{O}_i$ é aberto.

Aplicações sucessivas de (Top2) estabelecem que intersecções *finitas* de abertos são abertos.

Um *espaço topológico* é formado de um conjunto não vazio X juntamente com uma topologia τ em X . Formalmente, um espaço topológico é um par ordenado (X, τ) onde τ é topologia em X . Os elementos de X , independente da sua natureza, são chamados genericamente de *pontos*. Quando a topologia τ fica subentendida, nos referimos ao espaço topológico (X, τ) simplesmente como X .

Uma topologia se situa sempre entre dois extremos. Um extremo é dado pela *topologia trivial*, *indiscreta* ou *caótica* que consiste em tomar $\tau = \{X, \emptyset\}$, nenhuma topologia pode ter menos abertos que esta. No outro extremo temos a *topologia discreta* que consiste em tomar $\tau = \mathcal{P}(X)$, onde $\mathcal{P}(X)$ denota o conjunto das partes de X (conjunto composto de todos subconjuntos de X), nenhuma topologia pode ter mais abertos que esta.

Exemplo 1. A chamada *topologia padrão* é aquela definida no conjunto dos números reais. Um subconjunto $\mathcal{O}$ de $\mathbb{R}$ é dito aberto se é o vazio ou se para cada $x \in \mathcal{O}$ existe $\epsilon > 0$ tal

que $(x - \epsilon, x + \epsilon) \subset \mathcal{O}$. A coleção de todos os abertos de $\mathbb{R}$, assim definidos, constitui uma topologia em $\mathbb{R}$, denotada por $\tau_{\mathbb{R}}$.

Prova. i) $\mathbb{R}$ e $\emptyset$ são claramente abertos.

iii) Se $\mathcal{O}_\lambda$ é família qualquer de abertos então $\bigcup_\lambda \mathcal{O}_\lambda$ também é aberto. Se $x \in \bigcup_\lambda \mathcal{O}_\lambda$ então $x \in \mathcal{O}_{\lambda_0}$ para algum λ_0 , como $\mathcal{O}_{\lambda_0}$ é aberto, existe $(x - \epsilon, x + \epsilon) \subset \mathcal{O}_{\lambda_0}$ e, consequentemente, contido em $\bigcup_\lambda \mathcal{O}_\lambda$.

ii) Sejam $\mathcal{O}_1$ e $\mathcal{O}_2$ abertos e considere $\mathcal{O}_1 \cap \mathcal{O}_2$. Se $x \in \mathcal{O}_1 \cap \mathcal{O}_2$ então $x \in \mathcal{O}_1$ e $x \in \mathcal{O}_2$. Então existem ϵ_1 e ϵ_2 tal que $(x - \epsilon_1, x + \epsilon_1) \subset \mathcal{O}_1$ e $(x - \epsilon_2, x + \epsilon_2) \subset \mathcal{O}_2$. Tomando $\epsilon = \min(\epsilon_1, \epsilon_2)$, resulta que $(x - \epsilon, x + \epsilon) \subset \mathcal{O}_1 \cap \mathcal{O}_2$, de modo que $\mathcal{O}_1 \cap \mathcal{O}_2$ é aberto.

A *reta real* é o conjunto dos números reais com a topologia padrão. $\square$

Podemos também comparar duas topologias τ_1 e τ_2 definidas em X . Dizemos que τ_1 é *mais fina* que τ_2 se $\tau_2 \subset \tau_1$ e, ainda neste caso, dizemos que τ_2 é *mais grossa* que τ_1 . Ou seja, τ_1 é mais fina que τ_2 se tem mais abertos que τ_2 .

Se A é subconjunto de X , a *topologia induzida* em A , é a topologia cujos abertos são dados pela intersecção de A e os abertos $\mathcal{O}$ de X , ou seja,

$$\tau_A = \{A \cap \mathcal{O} \mid \mathcal{O} \in \tau\}.$$

Estes abertos de A respeitam as seguintes

- i) $\emptyset \cap A = \emptyset$, $X \cap A = A$.
- ii) $(\mathcal{O} \cap A) \cap (\mathcal{V} \cap A) = (\mathcal{O} \cap \mathcal{V}) \cap A$.
- iii) $\bigcup_{i \in I} (\mathcal{O}_i \cap A) = \left(\bigcup_{i \in I} \mathcal{O}_i \right) \cap A$.

Um subconjunto A de X junto com o topologia induzida τ_A é chamado *subespaço* de (X, τ) .

Sejam X não vazio e uma coleção arbitrária de subconjuntos U de X . A *topologia gerada* por U , denotada $\tau(U)$, é a intersecção de todas as topologias que contém U .

É um fato fácil de provar que a intersecção de topologias, $\tau_1 \cap \tau_2$, onde τ_1 e τ_2 são topologias em X , também é topologia.

A.1.2 Fecho, vizinhança e convergência

Dado um espaço topológico X , $p \in X$ é *ponto de acumulação* de $A \subset X$ se todo aberto que contém p também contém um ponto de A distinto de p . O conjunto dos pontos de acumulação de A chama-se conjunto *derivado* de A , denotado por A' . O *fecho* $\bar{A}$ de A é

o conjunto obtido unindo a A todos seus pontos de acumulação, $\bar{A} = A \cup A'$. Um conjunto A cujo fecho é igual ao espaço inteiro X é dito *denso* em X . O *interior* de A é a união de todos abertos $\mathcal{O} \subseteq A$, denotado por A° . A diferença $b(A) = \bar{A} - A^\circ$ é chamada *fronteira*, *borda* ou *contorno* de A . Antes de seguirmos para definição de fechados, vamos discutir sobre outras definições importantes para esta discussão.

Ainda sobre o espaço topológico X , $A \subset X$ é dito *vizinhança* do ponto $x \in X$ se existe um aberto $\mathcal{O}$ tal que $x \in \mathcal{O}$ e $\mathcal{O} \subset A$, se A é aberto diz-se que A é *vizinhança aberta* de x . Num espaço topológico a noção de convergência de sequências é introduzida de modo natural, sem depender de propriedades métricas de X . Dado o espaço topológico X , uma sequência (x_n) de pontos de X *converge* para $x \in X$ se para qualquer vizinhança aberta A de x existe $N \in \mathbb{N}$ tal que $x_n \in A$ para todo $n > N$.

Voltando a questão de fechados. Definimos F subconjunto *fechado* de um espaço topológico X se e somente se o seu complementar $F^c = X \setminus F$ é aberto. Seguem agora três teoremas sobre fechados cujas demonstrações constam nos textos de Szekeres para o primeiro e Lemos para os dois últimos.^{1,82}

Teorema 13. Num espaço topológico X , fecho de qualquer conjunto $A \subset X$ é um conjunto fechado. O interior A° é o maior aberto incluso em A . A borda $b(A)$ é um conjunto fechado.

Teorema 14. Num espaço topológico X , $F \subset X$ é fechado se e só se contém todos seus pontos de acumulação.

Teorema 15. Se F é fechado em um espaço topológico X , o limite de toda sequência convergente composta por elementos de F pertence a F .

A.1.3 Homeomorfismos

Sejam X e Y espaços topológicos. Uma aplicação $f : X \rightarrow Y$ é dita *contínua* se a pré-imagem $f^{-1}(B)$ de qualquer aberto $B \subset Y$ é um aberto de X . Ainda, se $f : X \rightarrow Y$ é aplicação bijetiva contínua cuja inversa f^{-1} também é contínua, diz-se que f é um *homeomorfismo* e que os espaços X e Y são *homeomorfos*, ou seja, indistinguíveis do ponto de vista topológico. Espaços homeomorfos são considerados apenas diferentes realizações de uma mesma estrutura subjacente a eles.

Propriedades preservados sob ação de homeomorfismos são chamadas *invariantes topológicos*, que podem ser desde números reais até propriedades estruturais como compacidade e conexidade. Pode-se entender topologia como o campo que vai caracterizar totalmente um espaço topológico através da descoberta de invariantes topológicos que cumpram esse objetivo.

A.1.4 Espaço de Hausdorff

Por topologias serem estruturas extremamente gerais podemos ser conduzidos a situações bem complicadas, como por exemplo a incapacidade de separarmos dois pontos distintos. Do ponto de vista físico isso é patológico, podemos ver uma sequência de pontos convergindo para dois pontos que representam estados/eventos fisicamente distintos porém matematicamente indistinguíveis.

Por isso, os chamados *axiomas de separação* costumam ser impostos sobre espaços topológicos. São eles

- i) Dados dois pontos de um espaço topológico X , pelo menos um deles pertence a um aberto que não contém o outro.
- ii) Dados dois pontos de um espaço topológico X , cada um deles pertence a um aberto que não contém o outro.
- iii) *Hausdorff* - Dados dois pontos de um espaço topológico X , existem abertos *disjuntos* cada um um contendo apenas um dos pontos.

Os axiomas acima são listados em ordem de força, de modo que iii) implica ii) e ii) implica i). Um *espaço de Hausdorff* é um espaço topológico que satisfaz o axioma iii).

Teorema 16. Num espaço Hausdorff uma sequência não pode convergir para dois limites distintos.

A prova deste teorema pode ser vista no texto de Lemos.⁸²

A.2 Tensores

A.2.1 Bases, componentes e mudança de base - componentes contravariantes e covariantes

Seja V espaço vetorial sobre um corpo $\mathbb{K}$, onde $\dim V = n$. Dada uma base $B = \{e_j\}_{j=1,\dots,n}$ de V um vetor $v \in V$ é escrito de maneira única como

$$v = \sum_{j=1}^n v^j e_j \equiv v^j e_j.$$

Esta é a chamada *convenção da soma de Einstein*, onde realizamos a soma sobre o conjunto de índices j que aparecem repetidos da forma acima (contraídos), assim omitindo a notação do somatório.

Todavia, sabemos também que V admite qualquer outra base $B' = \{e'_i\}_{i=1,\dots,n}$ de forma que as bases B e B' se relacionam por

$$\begin{aligned} e'_i &= e_j M^j_i \\ e_j &= e'_i N^i_j, \end{aligned}$$

onde $\mathbf{M} = [M^j_i]_{i,j=1,\dots,n}$ e $\mathbf{N} = [N^i_j]_{i,j=1,\dots,n}$ são as *matrizes de mudança e base* entre B e B' .

Como os vetores que compõem uma base são, por definição, linearmente independentes, temos que

$$\begin{aligned} e'_i &= e_j M^j_i = e'_k N^k_j M^j_i = e'_k \delta^k_i \\ &\rightarrow N^k_j M^j_i = \delta^k_i \\ &\rightarrow \mathbf{N}\mathbf{M} = \mathbf{1} \rightarrow \mathbf{N} = \mathbf{M}^{-1}. \end{aligned}$$

Outro fato imediato segue ao analisarmos a igualdade

$$v = v^j e_j = e'_k N^k_j v^j = v'^k e'_k,$$

ou seja, $v'_k = N^k_j v^j$. Matricialmente temos $(v)_{B'} = \mathbf{N}(v)_B$. Assim vemos que as componentes de um vetor $v \in V$ na mudança de base de $B \xrightarrow{\mathbf{M}} B'$ se alteram com a matriz inversa $\mathbf{N} = \mathbf{M}^{-1}$, *contravariam* com a mudança de base. Chamamos as componentes de $v \in V$ de *componentes contravariantes*.

Agora, seja V^* o espaço dual de V , formado pelas transformações lineares de V em $\mathbb{K}$, ou seja, $V^* \equiv \text{Lin}(V, \mathbb{K})$. Dizemos que uma base “natural” em V^* é dada por $B^* = \{e^i\}_{i=1,\dots,n}$, de forma que $e^i e_j = \delta^i_j$ para todo $e_j \in B$. Ora, se $c_i e^i = 0$, com c_i 's constantes arbitrárias

$$c_i e^i(e_j) = c_i \delta^i_j = c_j = 0,$$

ou seja, $c_i e^i = 0$ se e só se $c_i = 0$ para todo i . Seja $\omega \in V^*$ e $v \in V$

$$\omega(v) = \omega(v^i e_i) = v^i \omega(e_i) = \omega_i e^i(v),$$

portanto a base B^* produz todo o V^* . Além disso temos que $\dim V = \dim V^* = n$.

Elementos de V^* são chamados *funcionais lineares*, *covetores*, *vetores duais* ou *1-formas*. Ao atuarmos um covetor da base dual em um vetor $v \in V$ obtemos

$$e^i(v^j e_j) = v^j e^i e_j = \delta^i_j v^j = v^i.$$

Portanto, vemos que covetores da base coletam componentes contravariantes correspondentes ao seu índice.

Dado que a base direta B se transformou em B' segundo uma matriz $\mathbf{M}$, temos que a base dual B^* se transforma segundo a relação

$$e'^i = Q^i_k e^k,$$

onde agora $B'^* = \{e'^i\}_{i=1,\dots,n}$ preserva uma relação do tipo $e'^i e'_j = \delta^i_j$. Desta forma

$$\begin{aligned} \delta^i_j &= e'^i e'_j = Q^i_k e^k (e_l M^l_j) = Q^i_k M^l_j (e^k e_l) = Q^i_k M^k_j \\ &\rightarrow \mathbf{1} = \mathbf{Q}\mathbf{M} \rightarrow \mathbf{Q} = \mathbf{M}^{-1} = \mathbf{N}. \end{aligned}$$

Segue deste fato que, dado $\omega \in V^*$

$$\begin{aligned} \omega &= \omega'_i e'^i = \omega'_i Q^i_k e^k = \omega_k e^k \\ &\rightarrow \omega_k = \omega'_i Q^i_k \\ &\xrightarrow{M^k_j} \omega'_i Q^i_k M^k_j = \omega'_i \delta^i_j = \omega_k M^k_j, \end{aligned}$$

ou seja, $\omega'_j = \omega_k M^k_j$. Matricialmente temos $(\omega)_{B'^*} = (\omega)_{B^*} \mathbf{M}$. Assim vemos que as componentes de um covetor $\omega \in V^*$ na mudança de base de $B \xrightarrow{\mathbf{M}} B'$ se alteram com a própria matriz $\mathbf{M}$, *covariam* com a mudança de base. Chamamos as componentes de $\omega \in V^*$ de *componentes covariantes*.

Teorema 17. Para espaços vetoriais de dimensão finita, existe um isomorfismo entre V e V^{**} . Onde o bi-dual é definido como $V^{**} = \text{Lin}(V^*, \mathbb{K})$.

Prova. Para cada $v \in V$ fixo define-se $\tilde{v} : V^* \rightarrow \mathbb{K}$ por $\omega \mapsto \omega(v)$, associando um escalar a cada covetor.

Temos que $\tilde{v}$ é linear, já que, para $a, b \in \mathbb{K}$ e $\omega, \phi \in V^*$

$$\tilde{v}(a\omega + b\phi) = (a\omega + b\phi)(v) = a\omega(v) + b\phi(v) = a\tilde{v}(\omega) + b\tilde{v}(\phi).$$

Seja $\beta : V \rightarrow V^{**}$, definido por $v \mapsto \tilde{v}$, que também é linear, de modo que, para $a, b \in \mathbb{K}$, $u, v \in V$ e $\omega \in V^*$

$$\beta(au + bu)(\omega) = a\tilde{u}(\omega) + b\tilde{v}(\omega) = a[\beta(u)](\omega) + b[\beta(v)](\omega).$$

Sabemos que $B = \{e_i\}_{i=1,\dots,n}$ é base de V e $B^* = \{e^i\}_{i=1,\dots,n}$ é base de V^* , onde $e^i e_j = \delta^i_j$. Operando β em elementos das bases acima, temos

$$[\beta(e_j)](e^i) = \tilde{e}_j e^i = e^i(e_j) = \delta^i_j.$$

Vemos que $\{\beta(e_j) = \tilde{e}_j\}_{j=1,\dots,n}$ é base de V^{**} dual a base B^* , no sentido que definimos anteriormente. Assim, o mapa β é sobrejetor desde que cada $f \in V^{**}$ pode ser escrito na forma

$$f = u^i \beta(e_i) = \beta(u) = \tilde{u}, \quad u = u^i e_i.$$

Visto que $\{\beta(e_j) = \tilde{e}_j\}_{j=1,\dots,n}$ é base, então as componentes u^i (por consequência o vetor u) são unicamente determinados por f . Portanto β é um isomorfismo entre espaços vetoriais, independente da escolha de base. Por isso, no caso de dimensão finita, apenas V e V^* vão contar para nossa definição de tensores. $\square$

A.2.2 Tensor

Dados $V_1, \dots, V_N$ espaços vetoriais, agora sobre o corpo dos reais $\mathbb{R}$, uma *função multilinear* é tal que

$$T : V_1 \times \dots \times V_N \rightarrow \mathbb{R}$$

é linear em cada um dos seus argumentos separadamente.

As funções multilineares podem ser somadas e multiplicadas por escalares de maneira usual, de forma que, sendo $a, b \in \mathbb{R}$ e $v_1 \in V_1, \dots, v_N \in V_N$

$$(aT + bS)(v_1, \dots, v_N) = aT(v_1, \dots, v_N) + bS(v_1, \dots, v_N).$$

O espaço vetorial de todas as funções multilineares é denotado por $V_1^* \otimes \dots \otimes V_N^*$, chamado produto tensorial dos duais $V_1^*, \dots, V_N^*$.

Seja V um espaço vetorial de dimensão finita $\dim V = n$ sobre $\mathbb{R}$, fixamos $V_1 = \dots = V_r = V^*$ e $V_{r+1} = \dots = V_{r+s} = V$, onde agora $N = r + s$, nos referimos ao mapa

$$T : \underbrace{V^* \times \dots \times V^*}_r \times \underbrace{V \times \dots \times V}_s \rightarrow \mathbb{R}$$

como *tensor de posto* (r, s) em V , tal que o tensor é dito contravariante de ordem r e covariante de ordem s . O espaço vetorial de tais tensores é denotado como

$$T^{(r,s)}(V) = \underbrace{V \otimes \dots \otimes V}_r \otimes \underbrace{V^* \otimes \dots \otimes V^*}_s.$$

Dados tensores T de posto (r_1, s_1) e S de posto (r_2, s_2) chamamos de *produto tensorial* $T \otimes S$ o tensor de posto $(r_1 + r_2, s_1 + s_2)$ definido por

$$\begin{aligned} (T \otimes S)(\omega_1, \dots, \omega_{r_1}, \rho_1, \dots, \rho_{r_2}, v_1, \dots, v_{q_1}, u_1, \dots, u_{q_2}) \\ = T(\omega_1, \dots, \omega_{r_1}, v_1, \dots, v_{q_1}) S(\rho_1, \dots, \rho_{r_2}, u_1, \dots, u_{q_2}). \end{aligned}$$

A partir da definição, é direto verificar que o produto tensorial satisfaz

- i) $(R + S) \otimes T$ (distributivo).
- ii) $(R \otimes S) \otimes T = R \otimes (S \otimes T)$ (associativo).

Se $B = \{e_i\}_{i=1,\dots,n}$ é base de V e $B^* = \{e^i\}_{i=1,\dots,n}$ é base de V^* , dado $T \in T^{(r,s)}(V)$ podemos escreve-lo como

$$T = T^{i_1 \dots i_r}_{j_1 \dots j_s} e_{i_1} \otimes \dots \otimes e_{i_r} \otimes e^{j_1} \otimes \dots \otimes e^{j_s},$$

onde

$$T^{i_1 \dots i_r}_{j_1 \dots j_s} = T(e^{i_1}, \dots, e^{i_r}, e_{j_1}, \dots, e_{j_s}) := (T)^{i_1 \dots i_r}_{j_1 \dots j_s}$$

são as componentes do tensor T com respeito a base B de V .

$$e_{i_1} \otimes \dots \otimes e_{i_r} \otimes e^{j_1} \otimes \dots \otimes e^{j_s} \quad \text{com} \quad i_1, \dots, i_r, j_1, \dots, j_s = 1, \dots, n$$

formam uma base para $T^{(r,s)}(V)$, de forma que todo tensor T de posto (r, s) tem uma única expansão com componentes descritas acima.

Assim, temos que as componentes de uma combinação linear de tensores do mesmo posto, dado $a \in \mathbb{R}$, são dadas por

$$(T + aS)^{ij \dots}_{kl \dots} = T^{ij \dots}_{kl \dots} + aS^{ij \dots}_{kl \dots}$$

e as componentes do produto tensorial de T e S são dadas por

$$(T \otimes S)^{ij \dots pq \dots}_{kl \dots mn \dots} = T^{ij \dots}_{kl \dots} S^{pq \dots}_{mn \dots}.$$

Se realizarmos uma mudança de base $B \xrightarrow{\mathbf{M}} B'$, podemos ver como as componentes de T se transformam

$$\begin{aligned} T &= T^{i_1 \dots i_r}_{j_1 \dots j_s} e'_{i'_1} N^{i'_1}_{i_1} \otimes \dots \otimes e'_{i'_r} N^{i'_r}_{i_r} \otimes M^{j_1}_{j'_1} e'^{j'_1} \otimes \dots \otimes M^{j_s}_{j'_s} e'^{j'_s} \\ &= T'^{i'_1 \dots i'_r}_{j'_1 \dots j'_s} e'_{i'_1} \otimes \dots \otimes e'_{i'_r} \otimes e'^{j'_1} \otimes \dots \otimes e'^{j'_s}, \end{aligned}$$

desta forma temos que

$$T'^{i'_1 \dots i'_r}_{j'_1 \dots j'_s} = T^{i_1 \dots i_r}_{j_1 \dots j_s} N^{i'_1}_{i_1} \dots N^{i'_r}_{i_r} M^{j_1}_{j'_1} \dots M^{j_s}_{j'_s}.$$

A.2.3 Tensores simétricos e antissimétricos em $T^{(k,0)}(V)$ e $T^{(0,k)}(V)$

Antes de iniciarmos a discussão sobre simetrias de tensores precisaremos estabelecer alguns resultados sobre permutações, que aqui terão a prova de sua validade omitida com intuito de abreviar a questão. A prova dessas afirmações pode ser encontrada em livros de álgebra especializados como Lang e Artin, ou você pode encontrar os resultados em Szekeres.^{1,85,86}

De forma simplificada, *permutações* são bijeções que embaralham um conjunto finito de índices $(i_1, \dots, i_k) \mapsto (\sigma(i_1), \dots, \sigma(i_k))$. O conjunto de todas as $k!$ permutações forma o chamado *grupo simétrico* S_k . Denotamos por *transposição* uma permutação cíclica de tamanho 2, ou seja, uma permutação que mantém fixos $k-2$ elementos e inverte os outros 2 restantes. Afirmamos que, toda permutação pode ser escrita como uma série de composições de transposições. Afirmamos ainda que, a *paridade* do número de transposições que compõe uma permutação é invariante. Uma permutação é dita par(ímpar) se for expressa por composições de um número par(ímpar) de transposições. A *assinatura* ou *signal* de uma permutação σ , $\text{sgn}(\sigma) \equiv (-1)^\sigma$, é igual a 1 se ela for par ou -1 se ela for ímpar.

Um tensor $T \in T^{(k,0)}(V)$ é (totalmente) *simétrico* se, para k covetores de V^* e $\sigma \in S_k$, temos que

$$T(\omega_{\sigma(1)}, \dots, \omega_{\sigma(k)}) = T(\omega_1, \dots, \omega_k).$$

Dado um tensor qualquer $T \in T^{(k,0)}(V)$, o tensor definido como

$$T_{sim}(\omega_1, \dots, \omega_k) \equiv \frac{1}{k!} \sum_{\sigma \in S_k} T(\omega_{\sigma(1)}, \dots, \omega_{\sigma(k)})$$

é simétrico. Provar este fato é bastante simples, da definição acima de simetrização de tensores podemos escrever que

$$T_{sim}(\omega_{\sigma'(1)}, \dots, \omega_{\sigma'(k)}) = \frac{1}{k!} \sum_{\sigma \in S_k} T(\omega_{\sigma \circ \sigma'(1)}, \dots, \omega_{\sigma \circ \sigma'(k)}),$$

como S_k tem estrutura de grupo, então

$$T_{sim}(\omega_{\sigma'(1)}, \dots, \omega_{\sigma'(k)}) = \frac{1}{k!} \sum_{\pi \in S_k} T(\omega_{\pi(1)}, \dots, \omega_{\pi(k)}) = T_{sim}(\omega_1, \dots, \omega_k).$$

Toda a discussão acima é válida para o caso de $T^{(0,k)}(V)$.

Um tensor $T \in T^{(0,k)}(V)$ é (totalmente) *antissimétrico* se, para k vetores de V e $\sigma \in S_k$, temos que

$$T(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = (-1)^\sigma T(v_1, \dots, v_k).$$

Dado um tensor qualquer $T \in T^{(0,k)}(V)$, o tensor definido como

$$T_{skw}(v_1, \dots, v_k) \equiv \frac{1}{k!} \sum_{\sigma \in S_k} (-1)^\sigma T(v_{\sigma(1)}, \dots, v_{\sigma(k)})$$

é antissimétrico. A prova deste fato, tendo em mente a definição de tensor antissimétrico, é idêntica ao caso simétrico. O raciocínio também é válido para o caso de $T \in T^{(k,0)}(V)$.

Da definição de tensor antissimétrico seguem duas consequências interessantes, note que a segunda emana da primeira. Dado $T \in T^{(0,k)}(V)$, então

- i) $T(\dots, u, \dots, u, \dots) = 0$ para todo $u \in V$.
- ii) Se $\{u_i\}$ é um conjunto linearmente dependente de V , então $T(\{u_i\}) = 0$.

A.2.4 Sobe e desce de índices

Esta seção será dedicada a uma discussão sobre o fato do tensor métrico g_{ij} ter a famosa propriedade de subir e descer índices quando contraído com outros tensores.

Seja um espaço vetorial V sobre $\mathbb{R}$. Uma *forma não degenerada* (f.n.d) é uma função $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ que satisfaz as seguintes

- i) linear em ambos os argumentos (bilinear).

ii) $\langle u, v \rangle = \langle v, u \rangle$ (simétrica).

iii) Dado v e u em V , sendo $v \neq 0$, então $\langle u, v \rangle = 0$ implica $u = 0$ (não degenerada).

Se ainda, para todo v não nulo vale a propriedade adicional

iv) $\langle v, v \rangle > 0$ (positiva definida),

a f.n.d se torna um produto interno em V .

Seja $(V, \langle \cdot, \cdot \rangle)$ espaço vetorial munido com f.n.d real e base $B = \{e_i\}_{i=1, \dots, n}$. Agora, seja $\bar{g} : V \rightarrow V^*$, onde para todo $u, v \in V$, o mapa fica definido pela relação

$$\ll \bar{g}v, u \gg := \langle u, v \rangle = g(u, v).$$

Fixando v , sabemos que $\bar{g}(v) \in V^*$ e que a relação acima vale para qualquer $u \in V$. Então podemos dizer que, segundo a definição, a ação de $\bar{g}$ sobre um vetor é dada por

$$[\bar{g}(v)](\cdot) = \ll \bar{g}v, \cdot \gg = \langle \cdot, v \rangle = g(\cdot, v).$$

Ainda mais, é notável que $g(u, v) = \langle u, v \rangle$ é um mapa do tipo $g : V \times V \rightarrow \mathbb{R}$ que nada mais é que um tensor de posto $(0, 2)$ que herda as qualidades de simetria e não degenerescência pela identificação com a forma. Ele recebe o nome de *tensor métrico* da f.n.d. As componentes $g_{ij} = g(e_i, e_j) = \langle e_i, e_j \rangle$ da f.n.d são as componentes do tensor métrico com respeito a base B de V , onde

$$g = g_{ij}e^i \otimes e^j,$$

de forma que

$$g(u, v) = g_{ij}u^i v^j.$$

Atente que para todo vetor de V

$$\begin{aligned} \bar{g}(u) &= \bar{g}(v) \rightarrow g(u, \cdot) = g(v, \cdot) \\ &\rightarrow g(u - v, 0) = \langle u - v, 0 \rangle = 0 \\ &\rightarrow u = v. \end{aligned}$$

portanto $\bar{g}$ é injeção. Como V e V^* tem mesma dimensão, $\bar{g}$ é sobrejetora e portanto invertível.

A inversa $\bar{g}^{(-1)} : V^* \rightarrow V$, onde para todo $\omega, \rho \in V^*$, é definida pela relação

$$\ll \omega, \bar{g}^{(-1)}\rho \gg := g^{-1}(\omega, \rho)$$

e a ação de $\bar{g}^{(-1)}$ sobre um covetor é dada por

$$[\bar{g}^{(-1)}(\rho)](\cdot) = \ll \cdot, \bar{g}^{(-1)}\rho \gg = g^{-1}(\cdot, \rho),$$

de forma que $\bar{g} \circ \bar{g}^{(-1)} = \mathbf{1}_V$ e $\bar{g}^{(-1)} \circ \bar{g} = \mathbf{1}_{V^*}$.

Da simetria de g , para $\omega, \rho \in V^*$, temos que

$$\begin{aligned} g^{-1}(\omega, \rho) &= \ll \bar{g}(\bar{g}^{(-1)}\omega), \bar{g}^{(-1)}\rho \gg = g(\bar{g}^{(-1)}\omega, \bar{g}^{(-1)}\rho) = g(\bar{g}^{(-1)}\rho, \bar{g}^{(-1)}\omega) = \\ &= \ll \bar{g}(\bar{g}^{(-1)}\rho), \bar{g}^{(-1)}\omega \gg = \ll \rho, \bar{g}^{(-1)}\omega \gg = g^{-1}(\rho, \omega), \end{aligned}$$

ou seja, g^{-1} também é simétrica.

Analogamente, $g^{-1} : V^* \times V^* \rightarrow \mathbb{R}$ é um tensor simétrico não degenerado de posto $(2, 0)$ chamado de *tensor métrico inverso*, que com respeito a base B tem componentes $g^{ij} = g^{-1}(e^i, e^j)$, onde

$$g^{-1} = g^{ij}e_i \otimes e_j,$$

de forma que

$$g^{-1}(\omega, \rho) = g^{ij}\omega_i\rho_j.$$

Dado $v \in V$, podemos reescrever $\bar{g}(v) \in V^*$ como

$$\bar{g}(v) = \bar{g}(v^j e_j) = v^j [\bar{g}(e_j)],$$

então

$$\begin{aligned} (\bar{g}v)_i &= v^j [\bar{g}(e_j)](e_i) = v^j \ll \bar{g}e_j, e_i \gg = g(e_i, e_j)v^j = g_{ij}v^j \\ &= (v_k e^k)(e_i) = v_k \delta^k_i = v_i. \end{aligned}$$

Do mesmo modo, dado $\rho \in V^*$

$$\begin{aligned} (\bar{g}^{(-1)}\rho)^i &= \rho_j [\bar{g}^{(-1)}(e^j)](e^i) = \rho_j \ll e^i, \bar{g}^{(-1)}e^j \gg = g^{-1}(e^i, e^j)v^j = g^{ij}\rho_j \\ &= (\rho^k e_k)(e^i) = \rho^k \delta^i_k = \rho^i. \end{aligned}$$

Assim vemos a propriedade de subir e descer índices do tensor métrico aplicado nesse caso mais simples de vetores e covetores.

Outro fato direto dos nossos resultados é que

$$e_i = \bar{g}^{(-1)} \circ \bar{g}(e_i) = \bar{g}^{(-1)}(g_{ij}e^j) = g_{ij}g^{kj}e_k,$$

como o conjunto $\{e_i\}_{i=1,\dots,n}$ formam um conjunto linearmente independente segue que

$$g^{kj}g_{ji} = \delta^k_i$$

e as matrizes $[g_{ij}]$ e $[g^{ij}]$ são inversas uma da outra.

De maneira geral, a atuação $\bar{g}$ em qualquer uma das componentes contravariantes de um tensor geral T de posto (r, s) é uma função que leva $T^{(r,s)}(V) \rightarrow T^{(r-1,s+1)}(V)$, e analogamente, atuar $\bar{g}^{(-1)}$ em qualquer uma das componentes covariantes do mesmo tensor geral T é uma função que leva $T^{(r,s)}(V) \rightarrow T^{(r+1,s-1)}(V)$. Isso pode ser facilmente formalizado utilizando de todos os resultados acima e o produto tensorial.

APÊNDICE B – METROLOGIA

Metrologia é o ramo da ciência que busca estimar parâmetros de interesse via análise de dados obtidos experimentalmente. Em problemas de estimativa de parâmetros podemos obter informação sobre eles via amostragem de dados, sua distribuição de probabilidade e conhecimento sobre o processo físico envolvido. Algumas perguntas naturais podem surgir neste contexto, por exemplo:

-Quanta informação um conjunto de dados pode me fornecer sobre os parâmetros desconhecidos?

-Quão boa será a estimativa a partir dos dados obtidos?

Para responder estas perguntas devemos ver como uma modelagem matemático-estatística desenvolve resultados a cerca de um famoso quantificador de informação e sua relação com a qualidade e possível otimização da estimativa.

Neste apêndice queremos apresentar uma introdução à estimativa de um único parâmetro apresentando uma intuição para a definição da informação de Fisher e falar sobre a cota de Cramér-Rao como critério para a qualidade da estimativa, tudo isso sem fazer uso de conceitos abstrados de geometria, tornando a discussão mais próxima ao cenário físico usual. Para contexto clássico, seguiremos as notas de Zheng⁸⁷ e para o contexto quântico de estado puro a tese de Bastos.⁸⁸ Ao final da seção quântica, como âmbito de curiosidade, mostraremos como o emaranhamento pode interferir na estimativa.

Caso o leitor tenha maior interesse, recomendamos a leitura do livro de Cover & Thomas em teoria de informação clássica e as teses de Escher⁸⁹ e Taddei⁹⁰ em informação quântica.

Destaco que a notação empregada neste apêndice é diferente daquela vista, por exemplo, no Cap. 3. Mesmo assim, ainda é clara e de fácil compreensão.

B.1 Metrologia clássica

B.1.1 Contextualização

Valores de parâmetros desconhecidos que se desejam estimar ficam codificados nos resultados experimentais que, por sua vez, são codificados em variáveis estocásticas. A fim de estimar o valor real θ_v constante e desconhecido de um único parâmetro desejado é necessário um modelo matemático que utilize dos dados levando em conta o fenômeno físico envolvido e a distribuição estatística dos resultados. Portanto, no fundo, estimar é uma tarefa de característica teórica que opera sobre os resultados, dada sua estatística e a física enredada.

Podemos dividir o processo de estimar da seguinte forma:

de maneira que cada etapa é descrita abaixo,

- (1) Preparação do sistema físico (sonda) em uma configuração inicial.
- (2) Envio da sonda para o processo físico que insere a dependência no parâmetro de interesse.
- (3) Configuração final após processo físico.
- (4) Sonda submetida a uma medida experimental.

Repete-se (1)-(4) $n \gg 1$ vezes.

- (5) Estima-se o valor real do parâmetro utilizando-se das n medidas prévias.

As n repetições do experimento geram n medidas que serão agrupadas no conjunto $\mathbf{X} = (X_1, \dots, X_n)$, onde X_n é a *variável estocástica* associada a n -ésima rodada, discreta ou contínua.

Para construir a estimativa para θ_v , utilizamos um *estimador* $\hat{\theta}(\mathbf{X}) = r(X_1, \dots, X_n) = r(\mathbf{X})$ que é função aleatória dos dados experimentais com θ_{est} correspondendo a um dos valores possíveis assumidos por ele. Para o estimador é fundamental conhecer e ser possível tratar matematicamente o fenômeno físico dependente do parâmetro, além de ter acesso a distribuição de probabilidade de $\mathbf{X}$ consistente com o problema $f(\mathbf{X}, \theta_v)$, dado que $\theta_{est} = \theta_v$ foi impresso ao final de cada um dos n processos de medida. Entretanto isso não acontece em geral, além de que θ_v não é conhecido a princípio. Assim consideramos f em termos de valores θ que o parâmetro pode assumir em Θ , chamado *espaço de parâmetros*, de forma que agora $\hat{\theta}$ assume valores em Θ .

Idealmente $f(\mathbf{X}, \theta_v)$ só é alcançável para $n \rightarrow \infty$, ou seja, $n \rightarrow \infty$ medidas definem uma $f(\mathbf{X}, \theta_v)$ onde θ_v é um parâmetro determinado nesse limite. Na realidade, o que temos é um recorte, uma amostragem n finita associada a uma $f(\mathbf{X}, \theta)$ suposta densidade de probabilidade que se *assemelha* ao caso ideal. Queremos extrair de $f(\mathbf{X}, \theta)$ um valor que mais se aproxime do caso ideal. Fixo $\mathbf{X}$, denotamos $f(\mathbf{X}, \theta)$ por *função de máxima verossimilhança* (MLF) associada a distribuição de X , onde θ são os valores que o parâmetro pode assumir.

O chamado *estimador de máxima verossimilhança* (MLE)* é definido por

$$\hat{\theta}^{(ML)}(\mathbf{X}) = r^{(ML)}(\mathbf{X}) = \max_{\theta} [f(\mathbf{X}, \theta)] = \theta_{est}^{(ML)},$$

dado que a MLF supostamente semelhante ao caso ideal é determinada em termos dos dados experimentais e de θ , assim $\theta_{est}^{(ML)}$ deve ser uma estimativa razoável para o valor verdadeiro.

Devido a futura recorrência do uso, vamos apresentar também o logaritmo da MLF $l(\mathbf{X}, \theta) \equiv \log[f(\mathbf{X}, \theta)]$ (log-MLF), onde

$$\theta_{est}^{(ML)} = \max_{\theta} [l(\mathbf{X}, \theta)] = \hat{\theta}^{(ML)}(\mathbf{X}),$$

visto que a função logaritmo é monotonicamente crescente. Desta maneira temos uma outra maneira de extrair a estimativa $\theta_{est}^{(ML)}$ de $f(\mathbf{X}, \theta)$ olhando para derivas da log-MLF.

Derivadas de l em relação a θ indicam o quão f é sensível a variações do parâmetro. Denotando a ordem da derivada em relação ao parâmetro θ como linhas super escritas temos que

$$l'(\mathbf{X}, \theta) = \frac{f'(\mathbf{X}, \theta)}{f(\mathbf{X}, \theta)}.$$

Então quando θ assume valores próximos ao valor $\theta_{est}^{(ML)}$ a log-MLF deve ser próxima de zero,

$$l'(\mathbf{X}, \theta_{est}^{(ML)}) = \frac{1}{f(\mathbf{X}, \theta_{est}^{(ML)})} \left. \frac{\partial f(\mathbf{X}, \theta)}{\partial \theta} \right|_{\theta=\theta_{est}^{(ML)}} = 0.$$

Este processo para estimar θ_v é geral, já que não fizemos nenhuma hipótese sobre a correlação dos eventos. Ele nos deixa em circunstâncias de definir a famosa informação de Fisher.

B.1.2 Informação de Fisher

B.1.2.1 Caso de uma única observação

Considere a variável aleatória X para a qual a MLF é $f(X = x, \theta) := f(x, \theta)$, onde $\theta \in \Theta$ é parâmetro desconhecido. Intuitivamente, se um evento tem probabilidade pequena, então ocorrências desse evento nos trariam muita informação. Para uma variável aleatória X associada a MLF $f(x, \theta)$, se θ assume valores próximos ao valor verdadeiro do parâmetro implica que a MLF deve assumir um valor grande, ou equivalentemente, a derivada da log-MLF, $l'(x, \theta)$, nesse regime deve ser próxima de zero. Se $l'(x, \theta)$ é perto de zero, então é esperado que X não provê muita informação sobre θ ; por outro lado, $|l'(x, \theta)|$ ou $[l'(x, \theta)]^2$ é grande, então X provê muita informação sobre θ . Portanto podemos usar

*Vamos utilizar ao longo do texto a sigla ML, do inglês *most likelihood*, que significa máxima verossimilhança, assim como suas variações MLE –*most likelihood estimator*– e MLF –*most likelihood function*– que denotam, respectivamente, estimador de máxima verossimilhança e função de máxima verossimilhança.

$[l'(x, \theta)]^2$ para quantificar a informação provisionada por X . Como X é variável estocástica devemos considerar a média sobre os valores que θ pode assumir.

A *informação de Fisher*, para θ , contida na variável aleatória X é dada por

$$I(\theta) \equiv E_{\theta}\{[l'(x, \theta)]^2\} = \int [l'(x, \theta)]^2 f(x, \theta) dx.$$

Seguem resultados diretos das definições. Dado que $f(x, \theta)$ é suave e normalizada, $\int f(x, \theta) dx = 1$, é fácil provar que

$$E_{\theta}[l'(x, \theta)] = 0,$$

desta maneira a informação de Fisher é reescrita como

$$I(\theta) = \text{Var}_{\theta}[l'(x, \theta)].$$

Além disso, sabendo que a segunda derivada da log-MLF é dada por

$$l''(x, \theta) = \frac{f''(x, \theta)}{f(x, \theta)} - [l'(x, \theta)]^2,$$

temos que

$$I(\theta) = \text{Var}_{\theta}[l'(x, \theta)] = -E_{\theta}[l''(x, \theta)].$$

B.1.2.2 Transformação de parâmetro

Façamos X variável estocástica com MLF $f(x, \theta)$, onde θ é parâmetro desconhecido pertencente a Θ . $I_0(\theta)$ é a informação de Fisher em X . Suponha que o parâmetro θ é substituído por um parâmetro λ , onde $\theta = \varphi(\lambda)$ e φ é função suave. Façamos $I_1(\lambda)$ a informação de Fisher em X quando o parâmetro considerado é λ . Então

$$I_1(\lambda) = [\varphi'(\lambda)]^2 I_0[\varphi(\lambda)].$$

A prova deste fato segue de forma que, sendo $g(x, \lambda)$ a distribuição de probabilidade de X quando λ é considerado como parâmetro, temos que $g(x, \lambda) = f(x, \varphi(\lambda))$ e consequentemente $\log[g(x, \lambda)] = \log[f(x, \varphi(\lambda))] = l(x, \varphi(\lambda))$. Portanto

$$\begin{aligned} \frac{\partial}{\partial \lambda} \log[g(x, \lambda)] &= l'(x, \varphi(\lambda)) \varphi'(\lambda) \text{ e} \\ I_1(\lambda) &= E_{\lambda} \left\{ \left[\frac{\partial}{\partial \lambda} \log[g(x, \lambda)] \right]^2 \right\} = [\varphi'(\lambda)]^2 E_{\lambda} \{ [l'(x, \varphi(\lambda))]^2 \} = [\varphi'(\lambda)]^2 I_0(\varphi(\lambda)). \end{aligned}$$

B.1.2.3 Caso de n observações

Agora considere que temos um conjunto de n variáveis estocásticas $\mathbf{X} = (X_1, \dots, X_n)$ vindas de distribuições para qual a MLF é $f(x_i, \theta)$, com o $i = 1, \dots, n$ e valor de θ

desconhecido. Se as medidas são *independentes e igualmente distribuídas* (iid) então podemos escrever as seguintes

$$\begin{aligned} f_n(\mathbf{x}, \theta) &= \prod_{i=1}^n f(x_i, \theta), \\ l_n(\mathbf{x}, \theta) &= \log[f_n(\mathbf{x}, \theta)] = \sum_{i=1}^n \log[f(x_i, \theta)] = \sum_{i=1}^n l(x_i, \theta) \quad \text{e} \\ l'_n(\mathbf{x}, \theta) &= \frac{f'_n(\mathbf{x}, \theta)}{f_n(\mathbf{x}, \theta)}. \end{aligned}$$

A *informação de Fisher*, para θ , contida no conjunto de variáveis estocásticas $\mathbf{X}$ é dada por

$$I_n(\theta) \equiv E_\theta\{[l'_n(\mathbf{x}, \theta)]^2\} = \int [l'_n(\mathbf{x}, \theta)]^2 f_n(\mathbf{x}, \theta) d\mathbf{x}$$

Dada a suavidade e normalização de $f_n(\mathbf{x}, \theta)$, são válidas as seguintes

$$\begin{aligned} E_\theta[l'_n(\mathbf{x}, \theta)] &= 0 \quad \text{e} \\ I_n(\theta) &= \text{Var}_\theta[l'_n(\mathbf{x}, \theta)] = -E_\theta[l''_n(\mathbf{x}, \theta)], \end{aligned}$$

assim como no caso de uma única medida. Esses resultados são facilmente provados a partir das definições.

Note

$$I_n(\theta) = -E_\theta[l''_n(\mathbf{x}, \theta)] = -E_\theta\left[\sum_{i=1}^n l''(x_i, \theta)\right] = -\sum_{i=1}^n E_\theta[l''(x_i, \theta)] = nI(\theta),$$

trocando em palavras significa que a informação de Fisher de um conjunto aleatória iid de tamanho n é simplesmente n vezes a informação de uma única observação.

B.1.3 A cota de Cramér-Rao

Dado um conjunto $\mathbf{X} = (X_1, \dots, X_n)$ iid associado a MLF $f(x_i, \theta)$, com θ desconhecido, queremos examinar o quão boa é a estimativa que resulta de um estimador $\hat{\theta}$ do parâmetro determinando uma cota para sua variância. Afinal, em estatística a variância é uma grandeza matemática que expressa de forma razoável a qualidade da estimativa.

Façamos $\hat{\theta} = r(\mathbf{x})$ um estimador arbitrário de θ . Assuma que $E_\theta[\hat{\theta}] = m(\theta)$ e que a variância do estimador é finita. A covariância entre $\hat{\theta}$ e $l'_n(\mathbf{x}, \theta)$ é dada por

$$\begin{aligned} \text{Cov}_\theta[\hat{\theta}, l'_n(\mathbf{x}, \theta)] &= E_\theta\{(\hat{\theta} - E_\theta[\hat{\theta}])(l'_n(\mathbf{x}, \theta) - E_\theta[l'_n(\mathbf{x}, \theta)])\} \\ &= E_\theta\{[r(\mathbf{x})l'_n(\mathbf{x}, \theta)] - [m(\theta)l'_n(\mathbf{x}, \theta)]\} \\ &= \int r(\mathbf{x})f'_n(\mathbf{x}, \theta)d\mathbf{x} \\ &= \frac{\partial}{\partial\theta} \int r(\mathbf{x})f_n(\mathbf{x}, \theta)d\mathbf{x} = m'(\theta). \end{aligned}$$

Pela desigualdade de Cauchy-Schwarz

$$\{\text{Cov}_\theta[\hat{\theta}, l'_n(\mathbf{x}, \theta)]\}^2 \leq \text{Var}_\theta[\hat{\theta}] \text{Var}_\theta[l'_n(\mathbf{x}, \theta)]$$

obtemos a cota para a variância do estimador $\hat{\theta}$

$$\text{Var}_\theta[\hat{\theta}] \geq \frac{[m'(\theta)]^2}{nI(\theta)}.$$

Esta é a *desigualdade de Cramér-Rao*, ela mostra que conforme a informação cresce a variância do estimador diminui, ou seja, a qualidade da estimativa cresce. Por isso chamamos $I(\theta)$ de informação.

Um *estimador não viesado* ou *não viciado* é tal que $E_\theta[\hat{\theta}] = m(\theta) = \theta$, que implica $m'(\theta) = 1$. Em uma estimativa não viciada o valor médio do estimador coincide com um dos possíveis valores que θ pode assumir em Θ . Uma estimativa viciada ocorre quando todos os resultados são escalados e/ou deslocados ou sempre dão um mesmo valor independente da distribuição. Então, para um estimador não viesado a desigualdade de Cramér-Rao fica

$$\text{Var}_\theta[\hat{\theta}] \geq \frac{1}{nI(\theta)}. \quad (\text{B.1})$$

Chamamos o lado direito da desigualdade acima de *cota de Cramér-Rao* (CRB) que, sobre certas condições, nenhum estimador não viesado do parâmetro θ baseado em um conjunto iid de variáveis estocásticas de tamanho n terá variância menor que o CRB.

B.1.4 O teorema da distribuição assintótica do MLE

Considere o estimador $\hat{\theta}^{(ML)}$. Para uma amostragem de tamanho arbitrariamente grande, $n \rightarrow \infty$, temos que a distribuição do MLE se aproxima da distribuição normal com média θ_v e variância $(nI(\theta_v))^{-1}$. Como esse é um resultado limite, dizemos que o MLE é assintoticamente não viesado e nos referimos a variância do limite normal da distribuição como variância assintótica do MLE. Segue então o teorema que garante esse resultado.

Teorema 18. Seja $\mathbf{X} = (X_1, \dots, X_n)$ amostragem iid de tamanho n associada a distribuição $f(x, \theta)$ com θ desconhecido. Assuma que o valor verdadeiro de θ é θ_v e que estamos estimando via MLE. Então a distribuição de probabilidade de $\sqrt{nI(\theta_v)}(\hat{\theta}^{(ML)} - \theta_v) \sim \mathcal{N}(0, 1)$ quando $n \rightarrow \infty$. Em outras palavras, a distribuição assintótica do MLE é $\mathcal{N}(\theta_v, [nI(\theta_v)]^{-1})$

Nas notas de Zheng⁸⁷ existe uma prova deste teorema que vai utilizar resultados como a lei dos grandes números e o teorema do limite central, assuntos que estão muito além da discussão aqui pretendida. Recomendamos a leitura caso o leitor tenha interesse. O teorema acima indica que o limite assintótico de otimização do MLE alcança o CRB, por isso técnicas envolvendo MLE devem ser utilizadas com amostragens grandes.

B.2 Metrologia quântica de estado puro

B.2.1 Contextualização

Metrologia quântica vai estudar como técnicas quânticas permitem o ganho de vantagens em relação a abordagens clássicas. Apesar das diferenças entre física clássica e quântica, o processo geral para estimativa no caso quântico é similar ao que foi apresentado no caso clássico.

- (1) Sonda quântica preparada em um estado inicial descrito por ρ_0 .
- (2) Envio da sonda para o processo físico que insere a dependência no parâmetro de interesse.
- (3) Estado final após processo físico descrito por ρ_θ .
- (4) Sonda submetida a uma medida experimental, que é caracterizada pela ação de um conjunto de operadores no espaço de estados.

Repete-se (1)-(4) $n \gg 1$ vezes.

- (5) Estima-se o valor real do parâmetro utilizando-se das n medidas prévias.

As diferenças das etapas deste processo quântico geral de estimativa estão na caracterização do estado da sonda como estado quântico, em um protocolo quântico que insere a dependência no parâmetro e na definição de medida.

De fato, a principal diferença é notada na etapa (4). Seja o espaço de Hilbert $\mathcal{H}$, com $\dim \mathcal{H} = d$, e $\bar{\mathcal{S}}$ o conjunto convexo de operadores densidade ρ hermitianos, positivos semi-definidos de traço unitário em $\mathcal{H}$ que descrevem nossos estados quânticos. Medidas gerais em mecânica quântica são descritas por um outro conjunto de operadores $\{\pi_j\}$, chamados POVM's, que satisfazem para todo j

$$\pi_j \geq 0, \quad \pi_j = \pi_j^\dagger \quad e \quad \sum_j \pi_j = \mathbf{1}.$$

A estimativa depende da construção de um estimador $\hat{\theta}$. Como o processo de construção do estimador só depende dos resultados experimentais, esse passo é idêntico ao do caso clássico, portanto a precisão na estimativa, dado que $\hat{\theta}$ é não viesado, também vai ser dada por (B.1) para n repetições do experimento. Novamente, vemos que $I(\theta)$ limita a precisão máxima que se pode alcançar na estimativa do parâmetro, devemos revisitar a definição da informação de Fisher para que ela contenha os aspectos quânticos. Vale destacar que considera-se o estado inicial ρ_0 e o processo físico envolvidos conhecidos, desejamos estimar um único parâmetro codificado no estado final da sonda.

B.2.2 Cota de Cramér-Rao quântica

Seja o estado final da sonda dado por $\rho_\theta = \rho(\theta)$ para $\theta \in \Theta$ e um único aparato de medida experimental fixo representado pelo POVM $\{\pi_j\}$. O conjunto de probabilidades $p_j(\theta)$ de se obter um resultado j é dado por

$$p_j(\theta) = \text{Tr}(\rho_\theta \pi_j).$$

Dado este conjunto de probabilidades, pode-se calcular a informação de Fisher $I(\theta, \{\pi_j\})$ em analogia ao caso clássico, mas agora a quantidade de informação que o conjunto de dados pode me fornecer sobre o parâmetro depende de como faço a medida. Desta maneira, a maximização de $I(\theta, \{\pi_j\})$ sobre todas as possíveis medidas $\{\pi_j\}$ permitidas define a informação de Fisher quântica

$$I_q(\rho_\theta) \equiv \max_{\{\pi_j\}} [I(\theta, \{\pi_j\})] = \max_{\{\pi_j\}} \left\{ \sum_j \frac{1}{\text{Tr}(\rho_\theta \pi_j)} \left[\frac{\partial \text{Tr}(\rho_\theta \pi_j)}{\partial \theta} \right]^2 \right\}, \quad (\text{B.2})$$

onde a precisão da estimativa feita por um estimador $\hat{\theta}$ do parâmetro continua sendo dada pelo CRB

$$\text{Var}_\theta[\hat{\theta}] \geq \frac{1}{n I_q(\rho_\theta)}.$$

B.2.3 Informação de Fisher para estados puros e evoluções unitárias

Seja o estado inicial da sonda descrito por $\rho_0 = |\psi\rangle \langle \psi|$ puro, $\rho_0 = \rho_0^2$, onde $\langle \psi | \psi \rangle = 1$ e $|\psi\rangle$ representa um vetor de estado no espaço de Hilbert $\mathcal{H}$ na notação de Dirac. O processo físico dinâmico pelo qual a sonda irá passar será representado por um operador unitário $U_\theta = U(\theta) \equiv e^{-i\varphi(\theta)} V_\theta$ que codifica a informação sobre o parâmetro θ no estado final, sendo $V_\theta = V(\theta)$ operador unitário e $\varphi(\theta) = \int_0^\theta g(u) du$ função arbitrária de θ . Assim a atuação do processo físico sobre o estado inicial da sonda é dada por

$$U_\theta |\psi\rangle = e^{-i\varphi(\theta)} V_\theta |\psi\rangle := |\psi_\theta\rangle.$$

Sobre $|\psi_\theta\rangle$ é realizada a medida $\{\pi_j\}$, de forma que a probabilidade de resultado j ocorrer é

$$p_j(\theta) = \text{Tr}(|\psi_\theta\rangle \langle \psi_\theta | \pi_j).$$

Definindo o operador hermitiano

$$H_\theta = -i \frac{dV_\theta}{d\theta} V_\theta^\dagger$$

e sabendo que

$$\frac{d}{d\theta} |\psi_\theta\rangle = i \{H_\theta - g(\theta) \mathbf{1}\} |\psi_\theta\rangle$$

chegamos em

$$\frac{dp_j(\theta)}{d\theta} = \text{Tr} \left[\frac{d}{d\theta} (|\psi_\theta\rangle \langle \psi_\theta | \pi_j) \right] = -2 \text{Im} \{ \text{Tr}(\rho_\theta \pi_j \tilde{H}_\theta) \},$$

onde $\rho_\theta := |\psi_\theta\rangle\langle\psi_\theta|$ também é estado puro e $\tilde{H}_\theta := H_\theta - g(\theta)\mathbf{1}$.

É fato que $\text{Im}(z) \leq |z|^2$, para $z \in \mathbb{C}$, o que nos leva à

$$\left(\frac{dp_j(\theta)}{d\theta}\right)^2 \leq 4|\text{Tr}[(\sqrt{\rho_\theta}\sqrt{\pi_j})(\sqrt{\pi_j}\tilde{H}_\theta\sqrt{\rho_\theta})]|^2.$$

Pela desigualdade de Cauchy-Schwarz para o traço de matrizes positivo semi-definidas, $|\text{Tr}[A^\dagger B]|^2 \leq \text{Tr}[A^\dagger A]\text{Tr}[B^\dagger B]$, aplicada para $A = \sqrt{\rho_\theta}\sqrt{\pi_j}$ e $B = \sqrt{\pi_j}\tilde{H}_\theta\sqrt{\rho_\theta}$, obtemos

$$\left(\frac{dp_j(\theta)}{d\theta}\right)^2 \leq 4p_j(\theta)\text{Tr}[\tilde{H}_\theta\pi_j\tilde{H}_\theta\rho_\theta].$$

Portanto, da definição (B.2) temos que

$$\begin{aligned} I_q(\rho_\theta) &\leq \max_{\{\pi_j\}} \{4\text{Tr}[\tilde{H}_\theta^2\rho_\theta]\} \\ &\leq 4[\langle H_\theta^2 \rangle_{\psi_\theta} - 2g(\theta)\langle H_\theta \rangle_{\psi_\theta} + g^2(\theta)], \end{aligned}$$

onde $\langle A \rangle_{\psi_\theta} = \langle \psi_\theta | A | \psi_\theta \rangle$, com A operador hermitiano.

Como $g(\theta)$ é arbitrária, queremos um limite finito para a desigualdade acima, portanto vamos procurar por uma expressão de g que minimize a cota. A função g que minimiza a cota superior para a I_q é $g(\theta) = \langle H_\theta \rangle_{\psi_\theta}$, como visto na Prop. 9. Assim a cota para a informação fica

$$I_q(\rho_\theta) \leq 4\langle \Delta^2 H_\theta \rangle_{\psi_\theta},$$

onde $\langle \Delta^2 H_\theta \rangle_{\psi_\theta} := \langle H_\theta^2 \rangle_{\psi_\theta} - \langle H_\theta \rangle_{\psi_\theta}^2$ é a variância de H_θ no estado $|\psi_\theta\rangle$.

Agora vamos analisar as condições que saturam a cota acima para a informação e se essas condições são sempre satisfeitas. Lembre-se que utilizamos duas relações de desigualdade para chegar até aqui. Na primeira delas foi $\text{Im}(z) \leq |z|^2$, para $z \in \mathbb{C}$, a igualdade acontece quando $\text{Re}(z) = 0$, então a cota da informação é saturada quando $\text{Re}\{\text{Tr}[\rho_\theta\pi_j\tilde{H}_\theta]\} = 0$, ou equivalentemente

$$\text{Re}\{\langle \psi_\theta | \pi_j \tilde{H}_\theta | \psi_\theta \rangle\} = 0, \quad (\text{B.3})$$

lembrando que agora $g(\theta) = \langle H_\theta \rangle_{\psi_\theta}$.

A segunda foi a desigualdade de Cauchy-Schwarz, que é saturada quando $B = \alpha A$, para $\alpha \in \mathbb{C}$. A princípio, vamos fazer $\alpha = \alpha_j(\theta)$, lembrando que ρ_θ também é estado puro vemos que da relação $\text{Tr}(\sqrt{\pi_j}B) = \alpha_j(\theta)\text{Tr}(\sqrt{\pi_j}A)$ obtemos

$$\langle \psi_\theta | \pi_j \tilde{H}_\theta | \psi_\theta \rangle = \alpha_j(\theta) \underbrace{\langle \psi_\theta | \pi_j | \psi_\theta \rangle}_{\in \mathbb{R}}. \quad (\text{B.4})$$

Olhando para (B.4), se $\alpha_j(\theta)$ é imaginário puro então (B.3) é satisfeita, como $\alpha_j(\theta)$ é arbitrário, escolhe-lo imaginário puro faz as condições (B.3) e (B.4) equivalentes,

portanto é suficiente escolher apenas uma. Vamos escolher então (B.3). Definindo o estado normalizado

$$|\phi_\theta\rangle \equiv \frac{1}{\sqrt{\langle \Delta^2 H_\theta \rangle_{\psi_\theta}}} (H_\theta - \langle H_\theta \rangle_{\psi_\theta}) |\psi_\theta\rangle,$$

podemos reescrever a condição (B.3) como

$$\text{Re}\{\langle \psi_\theta | \pi_j | \phi_\theta \rangle\} = 0.$$

Assim, se agora $\pi_j = \pi_j(\theta)$, a condição é satisfeita escolhendo medidas projetivas $\pi_j(\theta) = |\psi_\theta\rangle \langle \psi_\theta|$ ou $\pi_j(\theta) = |\phi_\theta\rangle \langle \phi_\theta|$.

Como a informação de Fisher quântica envolve maximização sobre *todos* os POVM's possíveis, a existência de ao menos um que possibilita a saturação é suficiente para provar que

$$I_q(\rho_\theta) = 4 \langle \Delta^2 H_\theta \rangle_{\psi_\theta}$$

qualquer que seja a evolução unitária U_θ . Desta forma, dado um estimador $\hat{\theta}$ do parâmetro, o CRB para estados puros evoluindo unitariamente é escrito como

$$\text{Var}_\theta[\hat{\theta}] \geq \frac{1}{4n \langle \Delta^2 H_\theta \rangle_{\psi_\theta}},$$

que condiz com o resultado da equação (5.19) no contexto de uma única medida.

Portanto, um experimento realizado em sistemas quânticos no contexto de estados puros com dinâmica unitária, a precisão com que se pode estimar o valor de um parâmetro θ desconhecido não pode superar aquela descrita na cota acima.

B.2.4 Emaranhamento e o CRB

Ainda no caso de sistemas quânticos descritos por estados puros e evoluindo unitariamente, gostaríamos de apresentar como o emaranhamento pode interferir no processo de estimativa comparando com o caso em esse recurso quântico não é utilizado. Toda a ideia da montagem desse protocolo teórico de estimativa está descrito no texto de Maccone.⁹¹ O leitor interessado pode também conferir outros trabalhos em metrologia quântica da tríade Giovannetti, Lloyd e Maccone.^{92,93}

Nosso objetivo permanece sendo estimar um parâmetro desconhecido θ introduzido na sonda via operador unitário $U_\theta = \exp\{-i\theta H\}$, onde H gerador da dinâmica é hermitiano, conhecido e não dependente do parâmetro. A estratégia envolve n preparações paralelas de operações unitárias U_θ , ou seja, inicialmente teremos n sondas preparadas paralelamente que serão utilizadas em conjunto. Portanto, sendo $|\psi\rangle$ o estado inicial das n sondas, ele será transformado segundo a operação $U_\theta^{\otimes n} |\psi\rangle$, onde, $U_\theta^{\otimes n}$ é a transformação gerada por $h \equiv \sum_{j=1}^n H_j$ e H_j atua na j -ésima sonda. Sabemos pela seção anterior que qualquer que seja o esquema de medida empregado, para um estimador do parâmetro $\hat{\theta}$, $\text{Var}_\theta[\hat{\theta}]$ é

limitado inferiormente pelo CRB, $\text{Var}_\theta[\hat{\theta}]\Delta^2 h \geq 1/4\nu$, onde ν é o número de vezes que o processo de estimativa é repetido e $\Delta^2 h$ a variância de h no estado inicial $|\psi\rangle$ das sondas ($U_\theta^{\otimes n}$ comuta com h).

Sabemos também que o CRB é assintoticamente atingível no limite de $\nu \rightarrow \infty$, implicando que o mínimo erro $\text{Var}_\theta[\hat{\theta}]$ é obtido quando Δh é máximo. Trabalhando com erro quadrático médio $\delta\theta := \sqrt{\text{Var}_\theta[\hat{\theta}]}$, o CRB fica expresso como $\delta\theta \geq 1/[2(\nu\Delta^2 h)^{1/2}]$.

Sejam $|\lambda_M\rangle$ e $|\lambda_m\rangle$ autovetores de H_j correspondendo, respectivamente, ao maior e menor autovalores λ_M e λ_m . Se $|\psi\rangle$ é *separável*, o máximo valor de $\Delta^2 h$ acontece preparando cada sonda num estado tendo a superposição igualmente pesada de $|\lambda_M\rangle$ e $|\lambda_m\rangle$

$$\frac{1}{\sqrt{2}}(|\lambda_M\rangle + |\lambda_m\rangle).$$

Para estado separável $\Delta^2 h = \sum_{j=1}^n \Delta^2 H_j$, é facilmente calculável que $\Delta^2 H_j = (\lambda_M - \lambda_m)^2/4$. Desta forma o CRB fica

$$\delta\theta \geq 1/[\sqrt{n\nu}(\lambda_M - \lambda_m)]. \quad (\text{B.5})$$

Esta cota é saturada, por exemplo, por interferometria de Ramsey, isto é, preparando todas as sondas em $(|\lambda_M\rangle + |\lambda_m\rangle)/\sqrt{2}$ e medindo a probabilidade q cada sonda permaneça no mesmo estado na saída.

Agora se $|\psi\rangle$ é *emaranhado*, o valor máximo de $\Delta^2 h$ corresponde a um $|\psi\rangle$ onde temos uma superposição igualmente pesada de autovetores relativos ao maior e menor autovalores do gerador global h , ou seja, $n\lambda_M$ e $n\lambda_m$

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\lambda_M\rangle^{\otimes n} + |\lambda_m\rangle^{\otimes n}).$$

Utilizando que $(\sum_i a_i)^2 = \sum_i a_i^2 + 2\sum_{i<j} a_i a_j$ e que $\sum_{i<j}^n 1 = n!/[(n-2)!2!]$, temos que para o estado acima $\Delta^2 h = n^2(\lambda_M - \lambda_m)^2/4$. Desta forma o CRB fica

$$\delta\theta \geq 1/[\sqrt{\nu n}(\lambda_M - \lambda_m)]. \quad (\text{B.6})$$

Vemos que a cota no caso emaranhado produz uma melhora de um fator de $\sqrt{n}$ sobre a cota do caso separável (B.5).

A cota (B.6) é saturada utilizando o estado $(|\lambda_M\rangle^{\otimes n} + |\lambda_m\rangle^{\otimes n})/\sqrt{2}$ e estimando θ pelo observável $X \equiv |\lambda_m\rangle\langle\lambda_M| + |\lambda_M\rangle\langle\lambda_m|$ em cada sonda após a saída, note que $X^2 = \mathbf{1}$. Após um cálculo um tanto intrincado, tem-se que $\langle X^{\otimes n} \rangle_{out} = \cos^2[n\theta(\lambda_M - \lambda_m)]$ e conseqüentemente $(\langle \Delta^2 X^{\otimes n} \rangle_{out})^{1/2} = |\sin[n\theta(\lambda_M - \lambda_m)]|$. Depois de ν repetições do experimento o erro de θ pode ser obtido facilmente via propagação

$$\delta\theta = \frac{1}{\sqrt{\nu}}(\langle \Delta^2 X^{\otimes n} \rangle_{out})^{1/2} / \left| \frac{\partial \langle X^{\otimes n} \rangle_{out}}{\partial \theta} \right| = \frac{1}{\sqrt{\nu n}(\lambda_M - \lambda_m)}.$$

Este procedimento satura a cota (B.6) empregando apenas medidas separáveis, provando que medidas emaranhadas não são necessárias para se atingir a saturação.

APÊNDICE C – COTA DE CRAMÉR-RAO DO PONTO DE VISTA GEOMÉTRICO

Neste apêndice vamos apresentar geometricamente a derivação da cota de Cramér-Rao (CRB), percebendo que o conceito por trás do aparecimento da desigualdade é muito simples: dado um vetor que vive em um espaço qualquer, temos que sua norma sempre é maior ou igual a norma da sua componente que vive em um subespaço ortogonal. Este conceito emana naturalmente da estrutura geométrica de modelos estatísticos. Nossa discussão é encontrada, novamente, em diversas partes do texto de Amari,² por isso é importante ter em mente a notação apresentada no Cap. 3 para acompanhar a argumentação.

C.1 Famílias exponencial e mistura

Seja um modelo estatístico $\mathcal{P}(\mathcal{X}) \supset \mathcal{S} = \{p_\theta \mid \theta \in \Theta \subset \mathbb{R}^n\}$ onde as distribuições são expressas como

$$p_\theta = \exp \left[C(x) + \sum_{i=1}^n \theta^i F_i(x) - \chi(\theta) \right],$$

com $\{C, F_i\}_{i=1,\dots,n}$ funções de $\mathcal{X}$ e χ de Θ . Neste caso dizemos que o modelo é uma *família exponencial*. Da condição de normalização das distribuições temos que

$$\chi(\theta) = \log \int \exp \left[C(x) + \sum_{i=1}^n \theta^i F_i(x) \right] dx.$$

A n -upla $[\theta^i]$ será considerada sistema de coordenadas da família exponencial desde que $\{1, F_i\}_{i=1,\dots,n}$ sejam funções linearmente independentes, fazendo assim a parametrização $\theta \mapsto p_\theta$ biunívoca. Na nossa discussão sempre vamos assumir esta condição.

Para uma família exponencial, sendo $l_\theta = \log(p_\theta)$, temos que

$$\begin{aligned} \partial_i l_\theta &= F_i(x) - \partial_i \chi(\theta) \\ \partial_i \partial_j l_\theta &= -\partial_i \partial_j \chi(\theta), \end{aligned}$$

então, dada a definição das α -conexões em (3.6), temos que $\left(\Gamma_{kij}^{(\alpha=1)}\right)_\theta = -\partial_i \partial_j \chi(\theta) E_\theta[\partial_k l_\theta]$, que por sua vez é zero pela (3.4). Assim $[\theta^i]$ é sistema de coordenadas afim de $\mathcal{S}$, vamos então denotar $\nabla^{(e)} := \nabla^{(1)}$ chamada *e-conexão*.

Agora, expressando as distribuições por

$$p_\theta = C(x) + \sum_{i=1}^n \theta^i F_i(x),$$

onde $\{C, F_i\}_{i=1,\dots,n}$ são novamente funções de $\mathcal{X}$, dizemos que o modelo é uma *família mistura* com sistema de coordenadas $[\theta^i]$. Um exemplo de família mistura é a mistura

estatística de $n + 1$ distribuições de probabilidade $\{p_0, \dots, p_n\}$ com $\theta^i > 0$ e $\sum_i \theta^i < 1$, de forma que

$$p_\theta = \sum_{i=1}^n \theta^i p_i(x) + (1 - \sum_{i=1}^n \theta^i) p_0(x).$$

Para uma família mistura temos que

$$\partial_i l_\theta = \frac{F_i(x)}{p_\theta} \quad \text{e} \quad \partial_i \partial_j l_\theta = -\frac{F_i(x) F_j(x)}{p_\theta^2},$$

então $(\Gamma_{kij}^{(\alpha=-1)})_\theta = \partial_i \partial_j l_\theta + \partial_i l_\theta \partial_j l_\theta = 0$. Assim $[\theta^i]$ é sistema de coordenadas afim de $\mathcal{S}$, vamos então denotar $\nabla^{(m)} := \nabla^{(1)}$ chamada *m-conexão*.

C.2 Caso da variedade clássica

No caso de $\mathcal{X} = \{x_0, x_1, \dots, x_n\}$ finito e com $H = \{(\eta^1, \dots, \eta^n) \mid \eta^i > 0, \sum_{i=1}^n \eta^i < 1\}$, uma $p_\eta \in \mathcal{P}(\mathcal{X})$ pode ser expressa de maneira geral como

$$p_\eta = \begin{cases} \eta^i, & i \neq 0 \\ 1 - \sum_{i=1}^n \eta^i, & i = 0 \end{cases}.$$

isto faz com que $\mathcal{P}(\mathcal{X})$ em si seja um modelo estatístico formando uma variedade $(|\mathcal{X}| - 1)$ -dimensional, com $|\mathcal{X}|$ cardinalidade de $\mathcal{X}$. Em particular, $\mathcal{P}(\mathcal{X})$ é uma família mistura.

Um modelo $\mathcal{S} \subset \mathcal{P}(\mathcal{X})$ qualquer deve ser então subvariedade de $\mathcal{P}(\mathcal{X})$. Assim, como foi discutido na Sec. 3.3, a métrica de Fisher e as α -conexões em $\mathcal{S}$ são induzidas da estrutura afim do espaço de funções $\mathbb{R}^{\mathcal{X}} = \{A \mid A : \mathcal{X} \rightarrow \mathbb{R}\}$ onde $\mathcal{P}(\mathcal{X})$ está contido. No caso de $\mathcal{X}$ infinito, $\mathcal{P}(\mathcal{X})$ não é variedade, Amari faz uma breve discussão deste caso no fim da seção 2.5 de seu texto.²

Olhando para a geometria de $\mathcal{P}(\mathcal{X})$, vemos que ele é subconjunto de $\mathcal{A}_1 = \{A \mid \sum_x A(x) = 1\} \subset \mathbb{R}^{\mathcal{X}}$, assim $T_p(\mathcal{P})$ é naturalmente identificado com $\mathcal{A}_0 = \{A \mid \sum_x A(x) = 0\}$ (vide os teoremas da Sec. 3.6.2). Quando $X \in T_p(\mathcal{P})$ é elemento de $\mathcal{A}_0$, o denotamos por $X^{(m)}$, chamada *m-representação* de X , de maneira que

$$T_p^{(m)}(\mathcal{P}) \equiv \{X^{(m)} \mid X \in T_p(\mathcal{P})\} = \mathcal{A}_0.$$

Dado um sistema de coordenadas $[\eta^i]$ temos que um elemento da base natural de $X \in T_\eta(\mathcal{P})$ na *m-representação* será dado por $(\partial_i)^{(m)} = \partial_i p_\eta$, com $\partial_i = \partial / \partial \eta^i$.

Mergulhando $p \mapsto \log p$, identificando $\mathcal{P}$ com $\{\log p \mid p \in \mathcal{P}\} \subset \mathbb{R}^{\mathcal{X}}$, um vetor tangente $X \in T_p(\mathcal{P})$ é resultado da operação de X em $p \mapsto \log p$ e será denotado por $X^{(e)}$, chamada *e-representação* de X , onde

$$T_p^{(e)}(\mathcal{P}) \equiv \{X^{(e)} \mid X \in T_p(\mathcal{P})\} = \{A \in \mathbb{R}^{\mathcal{X}} \mid E_p[A] = 0\}.$$

Dado um sistema de coordenadas $[\eta^i]$, $(\partial_i)^{(e)} = \partial_i \log p_\eta$, desta forma

$$X^{(e)} = X^{(m)} / p(x). \tag{C.1}$$

Note que nessa representação, dado $X, Y \in T_p(\mathcal{P})$, a métrica de Fisher em $\mathcal{P}$ se escreve como

$$\langle X, Y \rangle_p = E_p[X^{(e)}Y^{(e)}].$$

Dada uma variedade Riemanniana $(\mathcal{S}, g = \langle \cdot, \cdot \rangle)$, para cada vetor tangente $X_p \in T_p(\mathcal{S})$, com $p \in \mathcal{S}$, temos um vetor cotangente $\omega_{X_p} \in T_p^*(\mathcal{S})$ tal que $\omega_{X_p} : Y_p \mapsto \langle X_p, Y_p \rangle_p$. A identificação $X_p \longleftrightarrow \omega_{X_p}$ estabelece um isomorfismo linear entre $T_p(\mathcal{S})$ e $T_p^*(\mathcal{S})$, de forma que o produto interno entre dois convetores é definido naturalmente como $\langle \omega_{X_p}, \omega_{Y_p} \rangle_p = \langle X_p, Y_p \rangle_p$.

Seja $f \in \mathcal{F}(\mathcal{S})$. Definimos $(df)_p \in T_p^*(\mathcal{S})$ por $(df)_p : X_p \mapsto X_p(f)$ chamada *diferencial* de f em p e $(\nabla f)_p \in T_p(\mathcal{S})$ o *gradiente* de f em p , de forma que

$$\langle (\nabla f)_p, X_p \rangle_p = (df)_p(X_p) = X_p(f), \quad \text{para todo } X_p \in T_p(\mathcal{S}).$$

Em termos de coordenadas $[\theta^i]$

$$\begin{aligned} (\nabla f)_\theta &= (\partial_i f)_\theta g^{ij}(\theta) (\partial_j)_\theta, \\ \|(df)_\theta\|_\theta^2 &= \|(\nabla f)_\theta\|_\theta^2 = (\partial_i f)_\theta (\partial_j f)_\theta g^{ij}(\theta). \end{aligned} \quad (\text{C.2})$$

Teorema 19. Dado $A : \mathcal{X} \rightarrow \mathbb{R}$ e a função $E[A]$ definida por $p \mapsto E_p[A]$ em $\mathcal{P}$, então

$$\text{Var}_p[A] = \|(dE[A])_p\|_p^2,$$

onde a norma é induzida da métrica de Fisher.

Prova. Dado $X \in T_p(\mathcal{P})$, temos que na e, m -representações do espaço tangente

$$\langle (\nabla E[A])_p, X \rangle_p = X(E[A]) = E_p[X^{(e)}A] = \sum_x X^{(m)}(x)A(x) = E_p[X^{(e)}(A - E_p[A])],$$

sabendo que $E_p[X^{(e)}] = 0$. Vemos que $(A - E_p[A])$ é vetor tangente de $\mathcal{P}$ e pela definição de $T_p^{(e)}(\mathcal{P})$ temos que $(\nabla E[A])_p = (A - E_p[A]) \equiv Y^{(e)} \in T_p^{(e)}(\mathcal{P})$. Então

$$\|(dE[A])_p\|_p^2 = \|Y^{(e)}\|_p^2 = E_p[(Y^{(e)})^2] = \text{Var}_p[A].$$

□

Teorema 20. Quando o domínio da função $E[A]$ é restrito ao um modelo estatístico $\mathcal{S} = \{p_\theta\} \subset \mathcal{P}$, então

$$\text{Var}_\theta[A] \geq \|(dE[A]|_{\mathcal{S}})_\theta\|_\theta^2,$$

onde a igualdade é válida se e só se

$$(A - E_\theta[A]) \in T_\theta^{(e)}(\mathcal{S}) \equiv \{X^{(e)} \mid X \in T_\theta(\mathcal{S})\}.$$

Prova. Pelo Teo. 19 e do fato que para qualquer $f \in \mathcal{F}(\mathcal{P})$ ao restringirmos o domínio da função $f|_{\mathcal{S}}$ o gradiente $(\nabla f|_{\mathcal{S}})_{\theta}$ será a projeção ortogonal de $(\nabla f)_p$ em $T_{\rho}(\mathcal{S})$, sendo a norma de uma projeção ortogonal sempre menor que a norma do vetor inteiro. $\square$

Agora a prova para o CRB é direta para $\mathcal{X}$ finito. Seja o estimador $A \equiv c_i \hat{\theta}^i(x)$ para um conjunto de valores fixos $(c_i \in \mathbb{R})_{i=1,\dots,n}$ de forma que $E_{\theta}[\hat{\theta}^i] = \theta^i$ para todo $\theta \in \Theta$. Pela relação (C.2) temos que

$$\begin{aligned} \left\| (\nabla E[c_k \hat{\theta}^k])_{\theta} \right\|_{\theta}^2 &= (\partial_i E[c_s \hat{\theta}^s])_{\theta} (\partial_j E[c_r \hat{\theta}^r])_{\theta} g^{ij}(\theta) = c_i c_j g^{ij}(\theta) \quad \text{e} \\ \text{Var}_{\theta}[c_k \hat{\theta}^k] &= c_i c_j E_{\theta}[(\hat{\theta}^i(x) - \theta^i)(\hat{\theta}^j(x) - \theta^j)] = c_i c_j v^{ij}(\theta), \end{aligned}$$

onde definimos $v^{ij}(\theta) \equiv E_{\theta}[(\hat{\theta}^i(x) - \theta^i)(\hat{\theta}^j(x) - \theta^j)]$, que são os componentes da chamada *matriz de covariância* $\mathbf{V}_{\theta}[\hat{\theta}]$, com $\hat{\theta} = [\hat{\theta}^1, \dots, \hat{\theta}^n] : \mathcal{X} \rightarrow \mathbb{R}^n$. Aplicando o Teo. 20, obtemos $c_i c_j v^{ij}(\theta) \geq c_i c_j g^{ij}(\theta)$, como os valores $(c_i)_{i=1,\dots,n}$ são quaisquer e fixos dado um estimador A , recuperamos o CRB

$$\mathbf{V}_{\theta}[\hat{\theta}] \geq \mathbf{G}^{-1}(\theta).$$

A condição necessária e suficiente para termos a saturação da cota é que $\mathcal{S}$ seja uma família exponencial (vide a discussão acerca da relação (5.7)), ou equivalentemente $(\hat{\theta}^i - \theta^i) \in T_{\theta}^{(e)}(\mathcal{S})$ para todo i .

C.3 Caso quântico

C.3.1 Covariâncias generalizadas

Os produtos internos $\{\ll \cdot, \cdot \gg_{\rho} \mid \rho \in \mathcal{S}\}$ em $\mathcal{A}$, conjunto de matrizes hermitianas, são chamados de *covariâncias generalizadas* se satisfazem as seguintes:

- i) $\ll \cdot, \cdot \gg_{\rho} \in \mathbb{R}$.
- ii) Tem dependência suave para todo $\rho \in \mathcal{S}$ e $A, B \in \mathcal{A}$.
- iii) Para todo $\rho \in \mathcal{S}$, $A, B \in \mathcal{A}$ e $U \in \mathcal{U}$ vale

$$\ll UAU^{\dagger}, UBU^{\dagger} \gg_{U\rho U^{\dagger}} = \ll A, B \gg_{\rho}.$$

- iv) Se $[\rho, A] = 0$ então

$$\ll A, B \gg_{\rho} = \text{Tr}(\rho AB).$$

Dois exemplos importantes de covariâncias generalizadas são o *produto interno simetrizado*

$$\ll A, B \gg_{\rho} = \frac{1}{2} \text{Tr}(\rho AB + \rho BA)$$

e o produto interno de Bogoliubov

$$\ll A, B \gg_\rho = \int_0^1 \text{Tr}(\rho^v A \rho^{1-v} B) dv$$

Fixo $\ll \cdot, \cdot \gg_\rho$, definimos a m -representação de um vetor tangente $X \in T_\rho$ como $X^{(m)} = [\Omega, \rho] \in T_\rho^u$, com Ω anti-hermitiano, e a e -representação como o operador $X^{(e)} \in \mathcal{A}$ tal que

$$\text{Tr}(X^{(m)} A) = \ll X^{(e)}, A \gg_\rho, \quad \text{para todo } A \in \mathcal{A}.$$

Se $X \in T_\rho^c$, então $[\rho, X] = 0$ e pela propriedade (iv) das covariâncias generalizadas

$$X^{(e)} = \rho^{-1} X^{(m)},$$

a correspondente quântica da relação (C.1).

Dado um sistema de coordenadas $[\theta^i]$ para $\mathcal{S}$ (ou uma subvariedade M de $\mathcal{S}$), analogamente, temos que $X \in T_\rho(\mathcal{S})$ na m -representação tem base natural dada por $(\partial_i)^{(m)} = \partial_i \rho_\theta$, já na e -representação vamos denotar a base por

$$(\partial_i)^{(e)} = L_i.$$

Quando o produto interno simetrizado é adotado como covariância generalizada em $\mathcal{S}$, para todo $A \in \mathcal{A}$, temos

$$\text{Tr}[(\partial_i)^{(m)} A] = \ll (\partial_i)^{(e)}, A \gg_\theta = \frac{1}{2} [\rho (\partial_i)^{(e)} A + \rho A (\partial_i)^{(e)}],$$

portanto

$$\partial_i \rho_\theta = \frac{1}{2} (\rho L_i + L_i \rho),$$

mostrando que L_i é a famosa *derivada logarítmica simétrica* introduzida na formulação quântica do CRB.^{94–97}

Adotando o produto de Bogoliubov como covariância generalizada, temos que

$$\partial_i \rho_\theta = \int_0^1 \rho_\theta^v L_i \rho_\theta^{1-v} dv.$$

Para $P = \rho_\theta$ e $Q = \rho_{(\theta+\theta^i)}$ com $\theta^i \ll 1$, utilizando a identidade

$$P - Q = \int_0^1 P^v (\log P - \log Q) Q^{1-v} dv,$$

válida para P, Q operadores positivos, temos que

$$L_i = \partial_i \log \rho_\theta.$$

Vale citar que, no texto de Ueda⁹⁸ ele relaciona as métricas de MCP (4.2) e covariâncias generalizadas definindo uma derivada logarítmica em termos das funções de MC. Ele ainda define um tipo de desigualdade de Cramér-Rao usando esses ingredientes.

C.3.2 As cotas quânticas

Seguem os teoremas que irão definir o CRB quântico:

Teorema 21. Seja A observável, $\mathcal{M}$ subvariedade de $\mathcal{S}$ e $\rho \in \mathcal{M}$. Temos que

$$\ll A - E_\rho[A], A - E_\rho[A] \gg_\rho \geq \|(dE[A]|_M)_\rho\|_\rho^2,$$

de forma que a igualdade é válida se e só se

$$A - E_\rho[A] \in T_\rho^{(e)}(\mathcal{M}) \equiv \{X^{(e)} \mid X \in T_\rho(\mathcal{M})\}.$$

Em particular, quando $\mathcal{M} = \mathcal{S}$ a igualdade é sempre válida.

Teorema 22. Seja $\mathcal{M} = \{p_\theta\}$ m -dimensional subvariedade de $\mathcal{S}$ com sistema de coordenadas $[\theta^i] \in \Theta \subset \mathbb{R}^m$. Se a m -upla $\mathbf{F} = (F^1, \dots, F^m) \in \mathcal{A}^m$ de observáveis no ponto $\theta \in \Theta$ é tal que

$$E_\theta[F^i] = \text{Tr}[\rho_\theta F^i] = \theta^i \quad \text{e} \quad \partial_j E_\theta[F^i] = \delta^i_j$$

para todo $i = 1, \dots, m$, então

$$\mathbf{C}_\theta[\mathbf{F}] \geq \mathbf{G}_\theta^{-1},$$

onde $\mathbf{G}_\theta$ é matriz com componentes g_{ij} de um tensor métrico g (vide as métricas (3.30)) com respeito ao sistema de coordenadas $[\theta^i]$ e $\mathbf{C}_\theta[\mathbf{F}]$ é a matriz de covariância generalizada com componentes

$$c_\theta^{ij} = \ll F^i - \theta^i, F^j - \theta^j \gg_\theta.$$

De um ponto de vista puramente matemático, estes teoremas são as extensões de seus análogos clássicos. Um ponto a se destacar é que a igualdade $\mathbf{C}_\theta[A] = \mathbf{V}_\theta[A]$ se e só se a covariância generalizada adotada for o produto interno simetrizado, em geral não é claro se $\mathbf{C}_\theta[A]$ terá algum significado estatístico dada uma covariância generalizada qualquer.

APÊNDICE D – MÉTRICA DE FUBINI-STUDY

Este apêndice é destinado a definirmos a métrica de Fubini-Study dada a estrutura suave de espaços projetivos. Esta métrica é a versão quântica da métrica de Fisher para o espaço de estados quânticos puros, visto a identificação natural deste com o espaço projetivo complexo.

D.1 Espaço projetivo

Seja V espaço vetorial sobre $\mathbb{K}$ de dimensão finita $\dim_{\mathbb{K}} = n + 1$, com $\mathbb{K} = \mathbb{R}, \mathbb{C}$, munido de uma forma não degenerada $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{K}$. Definimos uma relação de equivalência $\sim$ em V , de forma que, dados $u, v \in V$, $u \sim v$ se existe um $0 \neq a \in \mathbb{K}$ tal que $u = av$. Denotaremos a classe de equivalência de $v \in V$ como $[v]_{\sim}$.

Definimos o *espaço projetivo* como o quociente $\mathbb{P}_{\mathbb{K}}^n V \equiv V \setminus \{0\} / \sim$, onde

$$\mathbb{P}_{\mathbb{K}}^n V = \{[v]_{\sim} \mid v \in V\} = \{v \mid v = [v]_{\sim}, v \in V\}.$$

Cada classe de equivalência $[v]_{\sim}$ é uma reta passando pela origem que encontra a esfera $S_{\mathbb{K}}^n$ em dois pontos diametralmente opostos. A topologia de $\mathbb{P}_{\mathbb{K}}^n V$ é homeomorfa à esfera $S_{\mathbb{K}}^n$ com pontos diametralmente identificados.

Definimos também a chamada *aplicação quociente* que leva um elemento na sua classe de equivalência, $\pi : V \setminus \{0\} \rightarrow \mathbb{P}_{\mathbb{K}}^n V$ onde $v \rightarrow [v]_{\sim}$.

Proposição 22. O espaço projetivo real $\mathbb{P}_{\mathbb{R}}^n \equiv \mathbb{P}_{\mathbb{R}}^n(\mathbb{R}^{n+1})$ é uma variedade diferenciável.

Prova. Deveríamos primeiro mostrar que o espaço projetivo (sobre qualquer corpo $\mathbb{K}$) é espaço topológico Hausdorff com base enumerável. Porém, essa prova exige conceitos de topologia geral, como topologia quociente e bases topológicas, que não foram trabalhados neste texto no apêndice A. Para o leitor afeito a matemática, recomendamos a busca da prova deste fato no texto de Lee⁵¹ e a leitura sobre *topologias induzidas*.^{1,83}

Vamos achar então um atlas de cartas compatíveis que cobre $\mathbb{P}_{\mathbb{R}}^n$. Em $\mathbb{R}^{n+1}$ definimos a relação de equivalência $(x_1, \dots, x_{n+1}) \sim (x'_1, \dots, x'_{n+1})$ se existe um $\lambda \neq 0$ tal que $x_i = \lambda x'_i$ para todo $i = 1, \dots, n + 1$, assim $\mathbb{P}_{\mathbb{R}}^n$ é o conjunto de todas as classe de equivalência $[(x_1, \dots, x_{n+1})]_{\sim}$ que podem ser vistas como retas passando pela origem.

O conjunto de retas com $x_i \neq 0$, para $i = 1, \dots, n + 1$, denotado por $\mathcal{O}_i$, será um aberto na topologia induzida de $\mathbb{R}^{n+1}$ em $\mathbb{P}_{\mathbb{R}}^n$ via relação de equivalência, desta forma

$$\mathcal{O}_i \equiv \{[(x_1, \dots, x_{n+1})]_{\sim} \in \mathbb{P}_{\mathbb{R}}^n \mid x_i \neq 0\}.$$

Note que $\pi^{-1}\mathcal{O}_i = \mathbb{R}^{n+1} \setminus Hi$, onde $Hi = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_i = 0\}$ é o hiperplano com $x_i = 0$.

Definimos os mapas $\phi_i : \mathcal{O}_i \rightarrow \mathbb{R}^n$ por

$$\phi_i([(x_1, \dots, x_{n+1})]) = (y_{(i)}^1, \dots, y_{(i)}^n),$$

de maneira que, omitindo a razão $x_i/x_i = 1$, temos

$$y_{(i)}^1 = \frac{x_1}{x_i}, \dots, y_{(i)}^{i-1} = \frac{x_{i-1}}{x_i}, y_{(i)}^i = \frac{x_{i+1}}{x_i}, \dots, y_{(i)}^n = \frac{x_{n+1}}{x_i}.$$

A inversa $\phi^{-1} : \mathbb{R}^n \rightarrow \mathcal{O}_i$ pode ser vista como

$$\phi^{-1}(y_{(i)}^1, \dots, y_{(i)}^n) = [(x_1, \dots, x_{i-1}, 1, x_i, \dots, x_n)]_{\sim},$$

o que faz de ϕ uma bijeção.

Como $x_i = \lambda x'_i$, então $x'_j/x'_i = \lambda x_j/\lambda x_i = x_j/x_i$ para todo $j \neq i$, y^i é independentes da escolha do representante da classe $[(x_1, \dots, x_{n+1})]_{\sim}$. Uma vez que pelo menos um $x_i \neq 0$ em toda reta, cada representante esta contido em pelo menos um $\mathcal{O}_i$, assim o conjunto de cartas $(\mathcal{O}_i, \phi_i, y_{(i)}^j)$ cobre todo $\mathbb{P}_{\mathbb{R}}^n$.

Se $[(x_1, \dots, x_{n+1})]_{\sim} \subset \mathcal{O}_i \cap \mathcal{O}_j$, portanto $x_i \neq 0$ e $x_j \neq 0$, as funções de transição $\phi_j \circ \phi_i^{-1}$ são dadas por

$$y_{(j)}^k = \frac{x_k}{x_j} = \frac{x_k}{x_i} \frac{x_i}{x_j} = y_{(i)}^k y_{(j)}^i,$$

que são claramente invertíveis e suaves, desde que

$$y_{(i)}^k = y_{(j)}^k y_{(i)}^j.$$

Logo, esta é a estrutura diferenciável que faz de $\mathbb{P}_{\mathbb{R}}^n$ uma variedade diferenciável n -dimensional. $\square$

Proposição 23. O espaço projetivo complexo $\mathbb{P}_{\mathbb{C}}^n \equiv \mathbb{P}_{\mathbb{C}}^n(\mathbb{C}^{n+1})$ é uma variedade diferenciável de dimensão real $2n$.

Prova. Desde que $\mathbb{C} \simeq \mathbb{R}^2$, identificando $(x + iy) \mapsto (x, y)$, é direto que $\mathbb{C}^n \simeq \mathbb{R}^{2n}$, então podemos olhar para $\mathbb{C}^n$ como variedade de dimensão real $2n$.

Seguindo a mesma ideia da prova anterior, o aberto $\mathcal{O}_i$ é o conjunto de retas com $z_i \neq 0$, para $i = 1, \dots, n$. Definimos os mapas $\phi_i : \mathcal{O}_i \rightarrow \mathbb{C}^n$ por

$$\phi_i([(z_1, \dots, z_{n+1})]) = (w_{(i)}^1, \dots, w_{(i)}^n),$$

de maneira que, omitindo a razão $z_i/z_i = 1$, temos

$$w_{(i)}^1 = \frac{z_1}{z_i}, \dots, w_{(i)}^{i-1} = \frac{z_{i-1}}{z_i}, w_{(i)}^i = \frac{z_{i+1}}{z_i}, \dots, w_{(i)}^n = \frac{z_{n+1}}{z_i}.$$

Novamente w^i é independentes da escolha do representante da classe $[(z_1, \dots, z_{n+1})]_{\sim}$. Uma vez que pelo menos um $z_i \neq 0$ em toda reta, cada representante esta contido em pelo menos um $\mathcal{O}_i$, assim o conjunto de cartas $(\mathcal{O}_i, \phi_i, w_{(i)}^j)$ cobre todo $\mathbb{P}_{\mathbb{C}}^n$.

As funções de transição $\phi_j \circ \phi_i^{-1}$ são dadas por $w_{(j)}^k = w_{(i)}^k w_{(i)}^j$ que de novo são invertíveis e suaves. Portanto esta estrutura diferenciável faz de $\mathbb{P}_{\mathbb{C}}^n$ uma variedade diferenciável de dimensão real $2n$. $\square$

Proposição 24. O espaço tangente $T_{[p]_{\sim}}(\mathbb{P}_{\mathbb{K}}^n V)$, com $p \in V$, é isomorfo ao conjunto de transformações lineares da classe no perpendicular a classe $\text{Lin}([p]_{\sim}, p^{\perp})$.

Prova. Seja $\varphi \in \text{Lin}([p]_{\sim}, p^{\perp})$, dado $p \in V$, queremos provar que o mapa linear

$$\text{Lin}([p]_{\sim}, p^{\perp}) \xrightarrow{X} T_{[p]_{\sim}}(\mathbb{P}_{\mathbb{K}}^n V)$$

definido por

$$X_{\varphi} f \equiv \left. \frac{d}{dt} \right|_{t=0} \hat{f}(p + t\varphi(p))$$

é um isomorfismo, sendo

$$\begin{array}{ccc} V^{n+1} \setminus \{0\} & \xrightarrow{\pi} & \mathbb{P}_{\mathbb{K}}^n V \\ & \searrow \check{f} & \downarrow f \\ & & \mathbb{K} \end{array}$$

o diagrama que caracteriza as funções f e $\check{f} = f \circ \pi$. Note que função $\check{f}$ deve satisfazer a seguinte: $\check{f}(cp) = \check{f}(p)$ para todo $0 \neq c \in \mathbb{K}$.

Como $\dim \text{Lin}([p]_{\sim}, p^{\perp}) = \dim T_{[p]_{\sim}}(\mathbb{P}_{\mathbb{K}}^n V)$, basta provarmos a injetividade do mapa, ou seja, $\varphi \neq 0$ implica $t_{\varphi} \neq 0$. Já que X_{φ} deriva funções, basta achar uma $\check{f}$ tal que $X_{\varphi} f \neq 0$ se $\varphi \neq 0$. Vamos então construí-la explicitamente.

Conhecida uma transformação $\varphi \in \text{Lin}([p]_{\sim}, p^{\perp})$ não nula, para uma vizinhança aberta de $t \in \mathbb{K}$ em torno de zero, $\varphi(p)$ com $p \in V$ define uma reta $R \equiv (p + t\varphi(p))$ perpendicular a classe $[p]$. Por sua vez, a origem de V e a reta R definem um plano P . Ao restringirmos o domínio de $\check{f}$, sem perda de generalidade, ao plano furado $P^{\times} = P \setminus \{0\}$ denotaremos $\hat{f} \equiv \check{f} \circ \text{pr}_{P^{\times}}$. Para $s \in \mathbb{K} \setminus \{0\}$ temos uma família de retas $R_s \equiv s(p + t\varphi(p)) \in P^{\times}$ que “ligam” a origem de V a um ponto da reta R , em particular, quando $s = 1$ temos $R_{s=1} = R$. Fazemos então

$$\hat{f}\left(s(p + t\varphi(p))\right) = t,$$

fixando $s = 1$

$$\left. \frac{d}{dt} \right|_{t=0} \hat{f}(p + t\varphi(p)) = t_{\varphi} f = 1 \neq 0.$$

$\square$

Dado V sobre $\mathbb{K} = \mathbb{R}, \mathbb{C}$, com relação de equivalência $\sim$ e munido de forma não degenerada $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{K}$; estamos agora em condição de apresentar a métrica para $T_{[p]_{\sim}}(\mathbb{P}_{\mathbb{K}}^n V)$ que emana naturalmente da estrutura de V . Sejam duas transformações $\varphi_1, \varphi_2 \in \text{Lin}([p]_{\sim}, p^{\perp})$ e a classe $[p]_{\sim}$ de um ponto $p \in V$ definimos a *métrica de Fubini-Study* como

$$\ll t_{\varphi_1}, t_{\varphi_2} \gg_{[p]_{\sim}} \equiv \frac{\langle \varphi_1(p), \varphi_2(p) \rangle}{\langle p, p \rangle},$$

onde $t_{\varphi_1}, t_{\varphi_2} \in T_{[p]_{\sim}}(\mathbb{P}_{\mathbb{K}}^n V)$ e p é qualquer representante da classe $[p]_{\sim}$. Note que o denominador $\langle p, p \rangle$ faz com que a métrica seja independente da escolha do representante. Fubini-Study é a métrica mais simples que podemos construir dado a forma não degenerada que V possui.

Antes de partirmos para o caso da mecânica quântica, vamos mostrar outro fato útil. Dado $V = \mathbb{C}^{n+1}$ sobre $\mathbb{C}$ com $\dim_{\mathbb{C}} = n + 1$, colocando a relação de equivalência $\sim$ neste espaço ganhamos $\mathbb{P}_{\mathbb{C}}^n$. Agora seja a esfera complexa unitária

$$S_{\mathbb{C}}^n = \{z \mid z \in \mathbb{C} \text{ e } |z|^2 = 1\}$$

com uma relação de equivalência $\approx$, tal que $a \approx b$ se $a = e^{i\theta}b$, para $a, b \in S_{\mathbb{C}}^n$ e $\theta \in \mathbb{R}$. Denotaremos a classe de equivalência de $a \in S_{\mathbb{C}}^n$ como $[a]_{\approx}$.

Proposição 25. $\mathbb{P}_{\mathbb{C}}^n$ é difeomorfo a $S_{\mathbb{C}}^n / \approx$.

Prova. Seja o mapa

$$\mathbb{P}_{\mathbb{C}}^n \xrightarrow{\alpha} S_{\mathbb{C}}^n / \approx$$

definido por

$$[z]_{\sim} \mapsto z \mapsto \frac{z}{|z|} \mapsto \left[\left[\frac{z}{|z|} \right] \right]_{\approx}.$$

Este mapa é claramente uma aplicação suave, visto que quociente de função suave é suave.

A sobrejeção é direta. Queremos que toda classe em $S_{\mathbb{C}}^n / \approx$ tenha alguma classe em $\mathbb{P}_{\mathbb{C}}^n$ que é levada nela. Dada uma classe em $S_{\mathbb{C}}^n / \approx$ basta então tomarmos

$$\left[\left[\frac{z}{|z|} \right] \right]_{\approx} \mapsto [z]_{\sim} \xrightarrow{\alpha} \left[\left[\frac{z}{|z|} \right] \right]_{\approx}.$$

Para provar a injetividade do mapa tome duas classe $[z]_{\sim}, [w]_{\sim} \in \mathbb{P}_{\mathbb{C}}^n$ tal que $\alpha([z]_{\sim}) = \alpha([w]_{\sim})$, ou seja, $\left[\left[\frac{z}{|z|} \right] \right]_{\approx} = \left[\left[\frac{w}{|w|} \right] \right]_{\approx}$, desta forma

$$z = e^{i\theta} \frac{|z|}{|w|} w,$$

implicando que $[z]_{\sim} = [w]_{\sim}$. Portanto o mapa α é bijeção suave.

A inversa α^{-1} do mapa é dada também de forma direta por

$$\left[\left[\frac{z}{|z|} \right] \right]_{\approx} \rightarrow [z]_{\sim}$$

que também é claramente suave, portanto α é difeomorfismo. $\square$

Com este difeomorfismo fica claro que a $S_{\mathbb{C}}^n / \approx$ possui a mesma estrutura geométrica de $\mathbb{P}_{\mathbb{C}}^n$, portanto também vale Fubini-Study para $S_{\mathbb{C}}^n / \approx$.

D.2 Fubini-study no espaço de estados quânticos puros uniparamétricos

Seja a esfera unitária $S_{\mathbb{C}}^n$ do espaço de Hilbert $\mathcal{H}$ munido da forma hermitiana não degenerada $\langle \cdot | \cdot \rangle : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{C}$, de maneira que $\langle \psi | \psi \rangle = 1$ para todo $|\psi\rangle \in S_{\mathbb{C}}^n$. Definimos uma relação de equivalência $\sim$ em $S_{\mathbb{C}}^n$, de forma que, dados $|\psi\rangle, |\phi\rangle \in S_{\mathbb{C}}^n$, $|\psi\rangle \sim |\phi\rangle$ se $|\psi\rangle = e^{i\theta} |\phi\rangle$, com $\theta \in \mathbb{R}$. Aqui, denotaremos a classe de equivalência de $|\psi\rangle \in S_{\mathbb{C}}^n$ por $[[\psi]]$.

Sabemos que $S_{\mathbb{C}}^n / \sim$ tem estrutura de variedade do espaço projetivo complexo. Dada uma curva uniparamétrica V_t pertencente ao grupo unitário $\mathcal{U}$, com $t \in [0, \tau] \subset \mathbb{R}$, sua ação sobre uma $[[\psi]] \in S_{\mathbb{C}}^n / \sim$ gera uma curva $[[\psi_t]] = V_t[[\psi]]$ no espaço de projetivo. O vetor tangente em um ponto $[[\psi_t]]$ da curva será dado por

$$\frac{d}{dt}[[\psi_t]] = W_t[[\psi_t]], \quad \text{para todo } t \in [0, \tau],$$

identificando $W_t \equiv \dot{V}_t V_t^\dagger$.

Denotaremos $[[\psi_t]]$ como estado puro que descreve um sistema quântico em $t \in [0, \tau]$ e $[[\psi]]$ como estado inicial puro conhecido do sistema, afinal dinâmica unitária não interfere na pureza.

Teorema 23. Existe uma função real $f(t)$ e uma escolha suave de representantes para $[[\psi_t]]$ de forma que

$$|\psi_t\rangle = e^{i\theta(t)} V_t |\psi\rangle \quad \text{para todo } t,$$

onde $\theta(t) \equiv \int_0^t f(t') dt'$, tal que $\langle \psi_t | \dot{\psi}_t \rangle = 0$.

Prova. Para todo t , vamos pegar um representante da classe do estado inicial $e^{i\theta(t)} |\psi\rangle$ com $\theta(t)$ definido como no enunciado acima. Desta forma podemos escrever em termos de representantes

$$|\psi_t\rangle = e^{i\theta(t)} V_t |\psi\rangle = U_t |\psi\rangle,$$

com $U_t \equiv e^{i\theta(t)} V_t$.

A derivada de $|\psi_t\rangle$ é dada por

$$|\dot{\psi}_t\rangle = \Omega_t |\psi_t\rangle,$$

com $\Omega_t = \dot{U}_t U_t^\dagger = -i[H_t - f(t)]$ definindo $H_t = iW_t$.

Assim

$$\langle \psi_t | \dot{\psi}_t \rangle = 0 = -f(t) + \langle H_t \rangle_{\psi_t},$$

implica que $f(t) = \langle H_t \rangle_{\psi_t}$. □

Sabemos que $T_{[[\psi_t]]}(S_{\mathbb{C}}^n / \sim)$ é identificado com $\text{Lin}([\psi_t], |\psi^\perp\rangle)$, portanto existe uma única $\varphi \in \text{Lin}([\psi_t], |\psi^\perp\rangle)$ tal que, para qualquer representante de $[[\psi_t]]$, $\varphi(|\psi_t\rangle)$ é levado em $|\psi^\perp\rangle$. Identificando $|\dot{\psi}_t\rangle$ como um vetor perpendicular a toda classe $[[\psi_t]]$, de forma que $\varphi(|\psi_t\rangle) = |\dot{\psi}_t\rangle$, deve valer então que $\varphi = -i(H_t - \langle H_t \rangle_{\psi_t})$.

Portanto o coeficiente da métrica de Fubini-Study é dado por

$$g_t^{FS} = \left\langle \varphi(|\psi_t\rangle) \middle| \varphi(|\psi_t\rangle) \right\rangle = \text{Var}_{\psi_t}[H_t].$$

Este resultado concorda com a discussão da seção B.2.3, com a expressão (5.19) a menos de um fator de escala e com a Prop. 9. Novamente, parece que existe uma única fase tal que o $|\dot{\psi}_t\rangle$ é ortogonal em toda classe $[[\psi_t]]$.