Conexões duais

Fábio C. C. Meneghetti

IMECC — Unicamp

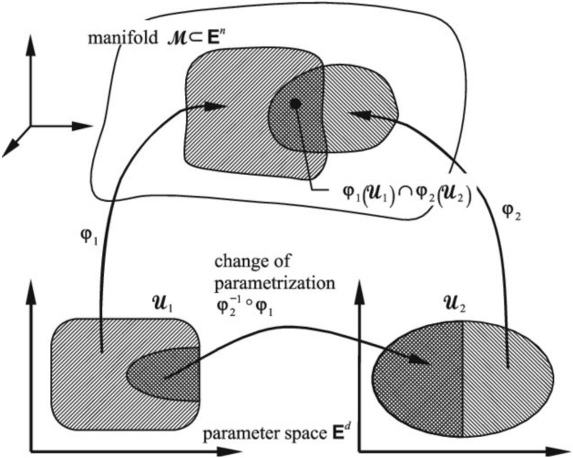
23 de julho de 2022

Variedades no espaço euclideano

- uma variedade de dimensão d em $\mathbb{R}^n$ é um conjunto $M \subset \mathbb{R}^n$ tal que:
  - para todo $p \in M$,
  - existem U aberto de $\mathbb{R}^n$, V aberto de $\mathbb{R}^k$ e
  - homeomorfismo $\varphi \colon U \cap M \to V$ (bijetora, contínua, e inversa contínua)
- ϕ é chamada **carta** e ϕ^{-1}, **parametrização**.
- uma coleção de cartas que cobre M é chamada **atlas**.

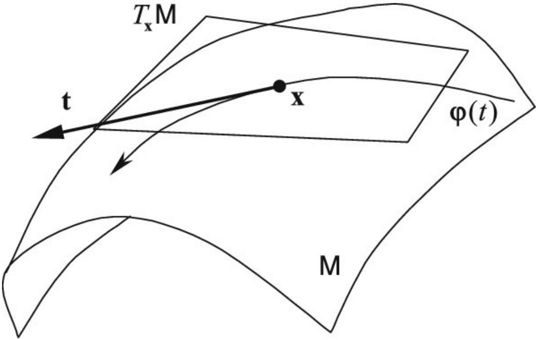
Variedades diferenciáveis

- estamos interessados nas variedades diferenciáveis:
  - (versão ingênua) pedimos que ϕ seja difeomorfismo (diferenciável, inversa diferenciável)
  - (versão abstrata) pedimos um atlas seja diferenciável, isto é, $\varphi^{-1} \circ \psi$ diferenciável para quaisquer cartas φ, ψ.
- **exemplos:**
  - esfera $\mathbb{S}^{n-1} = \left\{x \in \mathbb{R}^n : x_1^2 + \cdots + x_n^2 = 1\right\}$
  - qualquer cônica ou quádrica não-degenerada vistos em G.A.
  - gráficos de funções diferenciáveis


manifold $\mathcal{M} \subset \mathbf{E}^n$
$\varphi_1(\mathcal{U}_1) \cap \varphi_2(\mathcal{U}_2)$
φ_2
φ_1
change of
parametrization
$\varphi_2^{-1} \circ \varphi_1$
$\mathcal{U}_1$
$\mathcal{U}_2$
parameter space $\mathbf{E}^d$

Espaço tangente

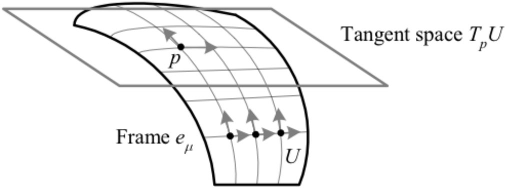
- cada ponto $p \in M$ tem um espaço tangente $T_p M$ (espaço vetorial de dimensão k)
- $v \in T_p M$ sse existe curva $\gamma\colon (-\varepsilon, \varepsilon) \to M$ tal que $\gamma(0) = p$ e $\gamma'(0) = v$
- a união disjunta $TM = \bigsqcup_{p \in M} T_p M$ é chamada **fibrado tangente**



$T_{\mathbf{x}}$M
t
x
φ(t)
M

Campos vetoriais

- um **campo vetorial** (diferenciável) em M é uma seção (diferenciável) de vetores $X\colon M \to TM$ tal que $X_p := X(p) \in T_pM$ para todo p. A família de todos os campos vetoriais é $\mathfrak{X}(M)$
- **campos canônicos:** no domínio de cada carta $\varphi\colon U \cap M \to V$ temos uma *base local de campos* $\partial_1, \ldots, \partial_d$, onde $\partial_i := \mathrm{d}\varphi^{-1} e_i$



Métrica riemanniana

- uma **métrica riemanniana** é um produto interno em cada espaço tangente de M,

$$g_p(V, W) = \langle V, W \rangle_p, \quad V, W \in T_pM,$$

  que é diferenciável, isto é, $\langle X_p, Y_p \rangle_p$ é diferenciável em p para quaisquer campos diferenciáveis X, Y.
- sendo produto interno, g precisa ser
  - simétrico ($\langle V, W \rangle = \langle W, V \rangle$),
  - positivo-definido ($\langle V, V \rangle \geq 0$ com $=$ sse $V = 0$) e
  - bilinear ($\langle aV + W, Z \rangle = a\langle V, Z \rangle + \langle W, Z \rangle$)

- um par (M, g) é chamado **variedade riemanniana**. Esses são os objetos de estudo da geometria riemannaiana
- o motivo é que, apenas definindo uma métrica, conseguimos estudar a geometria das variedades, em termos de:
    - linhas retas (geodésicas)
    - ângulos
    - curvatura
    - etc.
- **localmente** podemos representar uma métrica como uma matriz simétrica definida-positiva $g_{ij} = g(\partial_i, \partial_j)$ (para cada p)

Conexões afins

- é uma noção de derivada entre campos de vetores
- uma conexão afim é uma função $\nabla\colon \mathfrak{X}(M)^2 \to \mathfrak{X}(M)$, denotada $\nabla_X Y$, que satisfaz
  - $\nabla_{fX+Y} Z = f\nabla_X Z + \nabla_Y Z$ (linear na 1ª entrada com funções como coeficientes)
  - $\nabla_X(fY) = X(f) \cdot Y + f\nabla_X Y$ (satisfaz a regra de Leibniz na 2ª entrada)
- **exemplo:** se $\pi_p\colon \mathbb{R}^n \to T_pM$ é a projeção ortogonal no espaço T_pM, então $\nabla_X Y(p) := \pi_p(\mathrm{d}Y_p \cdot X_p)$ é uma conexão em M.

Teorema (Fundamental da geometria riemanniana)

Seja (M, g) variedade riemanniana. Então existe única conexão ∇^{LC} que é compatível com a métrica e livre de torção.

- ela é chamada conexão de Levi-Civita
- **compatível com a métrica:** significa que

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla_X Z\rangle.$$

- **livre de torção:** significa que o tensor torção $T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y]$ é constante igual a zero.
- para variedades no espaço euclideano, é a conexão mencionada no slide anterior

- localmente, toda conexão afim pode ser descrita em termos dos coeficientes — símbolos de Christoffel: Γ_{ij}^k ou $\Gamma_{ij,k}$:

$$\nabla_{\partial_i}\partial_j = \sum_{k=1}^{d} \Gamma_{ij}^k \partial_k,$$

$$\Gamma_{ij,k} = \langle \nabla_{\partial_i}\partial_j, \partial_k \rangle$$

- diremos que a conexão é *plana* em um ponto p se existe um sistema de coordenadas em torno de p no qual os símbolos de Christoffel se anulam.

Geometria da informação

- uma variedade estatística é uma família $M = \left\{ p_\theta : \theta \in \Theta \subset \mathbb{R}^d \right\}$ de funções densidade de probabilidade $p_\theta \colon \mathbb{R}^m \to \mathbb{R}_+$, que formam uma variedade diferenciável parametrizada por $\theta \mapsto p_\theta$
- podemos adicionar a métrica riemanniana da **informação de Fisher**, dada pela matriz:

$$g_{ij}(\theta) = \mathrm{E}_\theta \left[\partial_i \ell_\theta \partial_j \ell_\theta \right],$$

  onde $\ell_\theta(x) = \log p_\theta(x)$ e $\mathrm{E}_\theta[f] = \int_{\mathbb{R}^m} f(x) p_\theta(x)\, \mathrm{d}x$.
- a conexão de Levi-Civita ∇^{LC} tem coeficientes

$$\Gamma_{ij,k}(\theta) = \mathrm{E}_\theta \left[\left(\partial_i \partial_j \ell_\theta + \tfrac{1}{2} \partial_i \ell_\theta \partial_j \ell_\theta \right) \partial_k \ell_\theta \right]$$

Famílias exponenciais

- uma família exponencial é uma família de funções densidade de probabilidade $M = \{p_\theta : \theta \in \Theta\}$ da forma

$$p_\theta(x) = \exp\left(\sum_{i=1}^{d} \theta_i t_i(x) - \phi(\theta) + C(x)\right),$$

onde

  - $x \in \mathbb{R}^m$, $\theta \in \Theta$ um aberto de $\mathbb{R}^d$
  - $t_i : \mathbb{R}^m \to \mathbb{R}$ são funções chamadas *estatísticas suficientes*
  - ϕ é diferenciável e (pode-se mostrar que é necessariamente) convexa
  - C qualquer

- muitas distribuições importantes formam famílias exponenciais: normais uni/multivariadas, exponenciais, beta, gamma, Poisson, geométricas, etc.

- se a família é exponencial, a métrica e os símbolos de Christoffel têm expressões simples:

$$g_{ij}(\theta) = \partial_i \partial_j \phi(\theta), \qquad \Gamma_{ij,k} = \frac{1}{2} \partial_i \partial_j \partial_k \phi(\theta).$$

- uma das propriedades centrais encontradas por Amari, é que existe uma outra conexão, chamada **conexão exponencial** ∇^{e}, onde toda família exponencial é plana!

$$\Gamma^{\mathrm{e}}_{ij,k}(\theta) = \mathrm{E}_\theta \left[(\partial_i \partial_j \ell_\theta)(\partial_k \ell_\theta) \right],$$

$\Gamma^{\mathrm{e}}_{ij,k} \equiv 0$ numa família exponencial.

Famílias mistura

- outra família importante de distribuições são as famílias mistura, $\{p_\eta : \eta \in H\}$, dadas por

$$p_\eta(x) = C(x) + \sum_{i=1}^{d} \eta_i F_i(x),$$

onde

  - C é suave com $\int C(x)\,\mathrm{d}x = 1$.
  - F_i são suaves, L.I., com $\int F_i(x)\,\mathrm{d}x = 0$.

- **exemplo:** mistura de $p_1, \ldots, p_{d+1}$, dada por $p_\eta(x) = \eta_1 p_1(x) + \cdots + \eta_d p_d(x) + (1 - \sum_i \eta_i) p_{d+1}(x)$, $H = \left\{\eta \in [0,1]^d : \sum_i \eta_i \leq 1\right\}$

- $g_{ij}(\eta) = \mathsf{E}_\eta \left[\frac{F_i F_j}{p_\eta^2} \right]$, $\Gamma_{ij,k}(\eta) = -\frac{1}{2} \mathsf{E}_\eta \left[\frac{F_i F_j F_k}{p_\eta^3} \right]$
- nesse caso a informação de Fisher é o hessiano negativo da informação de Shannon $H(p_\eta) = \mathsf{E}_\eta[\ell_\eta]$:

$$g_{ij}(\eta) = -\partial_i \partial_j H(\eta)$$

- também existe uma **conexão mistura** ∇^{m}, na qual as famílias mistura são planas:

$$\Gamma^{\mathrm{m}}_{ij,k}(\eta) = \mathsf{E}_\eta \left[(\partial_i \partial_j \ell_\theta + \partial_i \ell_\theta \partial_j \ell_\theta)(\partial_k \ell_\theta) \right],$$

com $\Gamma^{\mathrm{m}}_{ij,k} \equiv 0$ se a família é mistura.

Conexões duais

- voltando para a abstração, duas conexões ∇, ∇^* são ditas **duais** se

$$X\langle Y, Z\rangle = \langle \nabla_X Y, Z\rangle + \langle Y, \nabla^*_X Z\rangle.$$

- (g, ∇, ∇^*) é chamada **estrutura dualística** para M

Proposição (Propriedades)

1. *o dual é único*
2. $\nabla^{**} = \nabla$
3. *se ambas são livres de torção, então* $\frac{1}{2}(\nabla + \nabla^*) = \nabla^{LC}$

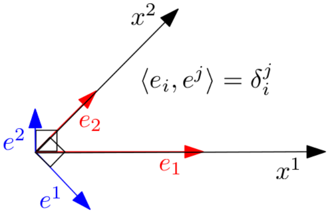
Teorema (Fundamental da geometria da informação)

∇ *é plana* $\iff$ ∇^* *é plana.*

- para qualquer variedade estatística, $(\nabla^{\mathrm{e}}, \nabla^{\mathrm{m}})$ é um par de conexões duais com relação à métrica de Fisher.
- em particular, para famílias exponenciais e mistura, $(g, \nabla^{\mathrm{e}}, \nabla^{\mathrm{m}})$ é uma estrutura **dualmente plana**.
- isso sugere que existem
  - sistema de coordenadas η, ∇^{m}-plano para uma família exponencial. De fato, $\eta = \operatorname{grad} \phi(\theta)$
  - sistema de coordenadas θ, ∇^{e}-plano para uma família mistura. De fato, $\theta = -\operatorname{grad} H(p_\eta)$

Bases duais

- se (M, g, ∇, ∇^*) é dualmente plana, então existem sistemas de coordenadas locais θ, η tais que
  - $\Gamma_{ij,k}(\theta) \equiv 0$,
  - $\Gamma_{ij,k}(\eta) \equiv 0$ e
  - se $\partial_i = \frac{\partial}{\partial \theta_i}$ e $\partial^j = \frac{\partial}{\partial \eta_j}$ então $\langle \partial_i, \partial^j \rangle = \delta_{ij}$ (as bases são duais)



Divergência de Kullback-Leibler

- é uma medida de divergência entre distribuições de probabilidade:

$$D_{\mathrm{KL}}(p\|q) := \mathrm{E}_p\left[\log\frac{p}{q}\right]$$

- $D_{\mathrm{KL}}(p\|q) = 0 \iff p = q$ em quase todo ponto.
- é *assimétrica*, $D_{\mathrm{KL}}(p\|q) \neq D_{\mathrm{KL}}(q\|p)$
    - essa assimetria tem a ver com a dualidade (∇, ∇^*)

Geodésicas duais

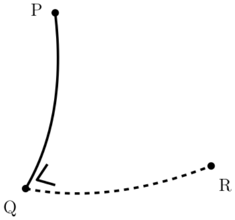
- cada conexão ∇ define uma noção de geodésica: curvas γ tais que $\nabla_{\dot{\gamma}}\dot{\gamma} \equiv 0$
- assim, em uma variedade estatística temos 3 noções de geodésicas:
  - ∇^{LC}-geodésicas,
  - ∇^{e}-geodésicas,
  - ∇^{m}-geodésicas

Pitágoras

Teorema (Pitágoras generalizado)

Sejam $p, q, r \in M$, tais que as geodésicas γ_{pq} e γ^*_{qr} são ortogonais. Então

$$D_{KL}(p\|q) + D_{KL}(q\|r) = D_{KL}(p\|r)$$



Projeção informacional

- sejam $S \subset M$ subvariedade, $p \in M$. A **projeção informacional** de p em S é o ponto $p^* \in S$ que minimiza $D_{\mathrm{KL}}(p^* \| p)$.

Teorema (Projeção)

Se S é ∇^m-geodésica, então a ∇^e-geodésica ligando p e p^* é ortogonal a S.

Teorema (Esfera de KL)

*Sejam $p \in M$ e $S_r = \{q \in M : D_{KL}(q \| p) = r\}$. Então toda ∇^e-geodésica passando por p intersecta S_r **ortogonalmente**.*

Referências

1. Geometria diferencial.
    - Manfredo P. do Carmo — *Geometria diferencial de curvas e superfícies*
    - Manfredo P. do Carmo — *Geometria riemanniana*
2. Geometria da informação.
    - Amari, Nagaoka — *Methods of information geometry*
    - Calin, Udriște — *Geometric modeling in probability and statistics*
    - Nielsen — *Elementary introduction to information geometry*