

最適化, 統計, 機械学習におけるホロノミック法

高山信毅*(神戸大学, 数学科)

1. “ホロノミック (holonomic)” とは?
2. 力学系としてのホロノミック勾配法 (holonomic gradient method, HGM) および最尤推定 (maximal likelihood estimation, MLE).
3. 最適化, 統計, 機械学習への応用.

代数幾何における次元

$I : K[x] := K[x_1, \dots, x_n]$. $K = \mathbb{C}$ のイデアル.

$\text{ord}_u x^\alpha := \langle u, \alpha \rangle$, $x^\alpha := \prod_{i=1}^n x_i^{\alpha_i}$

$F_k : \bigoplus_{\text{ord}_1(x^\alpha) \leq k} Kx^\alpha$, $\mathbf{1} := (1, \dots, 1)$

I の **ヒルベルト多項式** : $H(k) = \dim_K \frac{F_k}{F_k \cap I}$.

例: $n = 2$, $I = \langle x_1 x_2 \rangle$.

$F_k / (F_k \cap I) = K + Kx_1 + \dots + Kx_1^k + Kx_2 + \dots + Kx_2^k$ の時

$H(k) = 2k + 1$.

$H(k)$ の k についての**次数** を I の **次元** と定義する. $V(I)$ が複素多様体 (manifold) ならこれは複素多様体としての次元と一致.

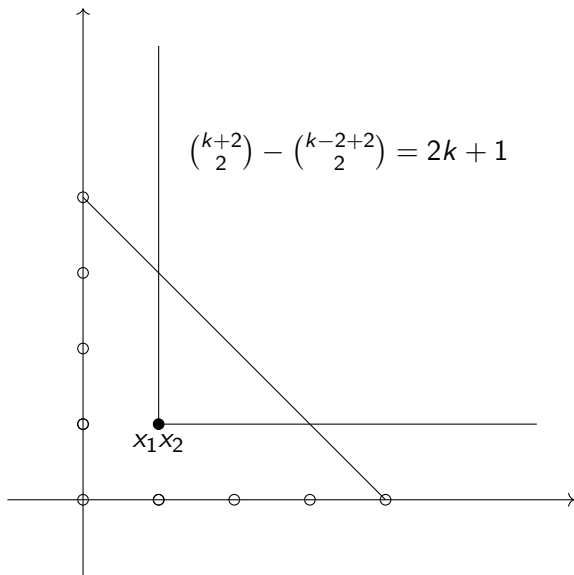


Figure: $\frac{F_4}{F_4 \cap I}$, $I = \langle x_1x_2 \rangle$

ワイル代数 (多項式係数線形偏微分作用素の環)

$n = 1, x = x_1.$

$D_1 = K\langle x, \partial_x \rangle, \partial_x x = x\partial_x + 1.$

$I : D_1$ の左イデアルとは (1) $L, M \in I$ なら $L - M \in I$. (2) $L \in I$, なら任意の $M \in D_1$ に対して $ML \in I$

D_1 の任意のイデアルは有限個の元 $L_1, \dots, L_m$ で生成される. グレブナー基底の理論がなりたつ.

f : 関数. ∂_x や x の f への作用は $\partial_x \bullet f = \frac{df}{dx}$ および $x \bullet f = xf$. 関数 f が $L_1, \dots, L_m$ により零化される (annihilated) なら, f は I の任意の元で零化される.

例: $I = \langle x\partial_x - 1, \partial_x^2 \rangle$. $f = x$ は I で零化される.

D_1 の任意の左イデアル $I (\neq 0, D_1)$ は ホロノミックイデアル.

多変数の場合 “ホロノミック” イデアルをどう定義するか?

ワイル代数: $D_n = K\langle x_1, \dots, x_n, \partial_1, \dots, \partial_n \rangle$, $x_i x_j = x_j x_i$,
 $\partial_i \partial_j = \partial_j \partial_i$, $\partial_i x_j = x_j \partial_i + \delta_{ij}$ ($\partial_i = \frac{\partial}{\partial x_i}$)

$u, v \in \mathbb{R}^n$, $\text{ord}_{(u,v)}(x^\alpha \partial^\beta) := \langle u, \alpha \rangle + \langle v, \beta \rangle$.

定義 (I.N.Bernstein=J.Bernstein, 1972).

$F_k = \bigoplus_{\text{ord}_{(1,1)}(x^\alpha \partial^\beta) \leq k} K x^\alpha \partial^\beta$. I を D_n の左イデアルとする.

$H(k) = \dim_K \frac{F_k}{F_k \cap I}$. ヒルベルト多項式 $H(k)$ の次数が n の時 I を
ホロノミックイデアル と定義する.

事実: ホロノミックイデアルは I はすべての変数 x_i について以下の形の常微分作用素を含む

$$\sum_{k=0}^{r_i} s_{ik}(x) \partial_i^k, \quad s_{ik}(x) \in K[x], \quad i = 1, \dots, n \quad (1)$$

証明. K -線形写像

$$p : F_k \cap K\langle x_1, \dots, x_n, \partial_i \rangle \longrightarrow \frac{F_k}{F_k \cap I}$$

左辺の次元は $\binom{k+(n+1)}{n+1} = O(k^{n+1})$. よって $p^{-1}([0])$ は 0 でない元を含む. $\square$

上述の形の常微分作用素が生成する左イデアルはホロノミックとは限らない. 例: $n = 2$.

$I = \langle L_1 := (x_1^3 - x_2^2)\partial_1 + 3x_1^2, L_2 := (x_1^3 - x_2^2)\partial_2 - 2x_2 \rangle$
($I \bullet (x_1^3 - x_2^2)^{-1} = 0$). $H(k) = \frac{k^3}{2} + 2k^2 + \frac{k}{2} + 2$. したがって I はホロノミックイデアルでない. ちなみに I に $2x_1\partial_1 + 3x_2\partial_2 + 6$ を加えるとホロノミックになる.

R_n : 有理式係数の線形偏微分作用素の環 (有理ワイル代数)

$R_n = \mathbb{C}(x)\langle \partial_1, \dots, \partial_n \rangle$. J : 上述の (1) なる形の元で生成される R_n の左イデアル.

事実: $R_n J \cap D_n$ はホロノミックイデアル.

例: $\frac{2}{p}x_1L_1 + \frac{3}{p}x_2L_2 = 2x_1\partial_1 + 3x_2\partial_2 + 6, p = x_1^3 - x_2^2$.

Theorem 1 (I.N.Bernstein, 1972¹)

1. D_n の左イデアル $I \subsetneq D_n$ のヒルベルト多項式の次数は n 以上.
2. I がホロノミックイデアルなら $(I + x_n D_n) \cap D_{n-1}$ (制限イデアル, *restriction ideal*) および $(I + \partial_n D_n) \cap D_{n-1}$ (積分イデアル, *integration ideal*) は D_{n-1} のホロノミックイデアル.

事実: 急減少関数 f がホロノミックイデアル $I \subset D_n$ で零化されるなら, $n-1$ 変数 x' の関数 $g(x') := \int_{-\infty}^{\infty} f(x) dx_n$ は積分イデアルで零化される.

証明. $L = L_1 + \partial_n L_2 \in (I + \partial_n D_n) \cap D_{n-1}$. よって

$$\begin{aligned} L \bullet g(x') &= \int_{-\infty}^{\infty} L_1 \bullet f dx_n + \int_{-\infty}^{\infty} \partial_n L_2 \bullet f dx_n = \int_{-\infty}^{\infty} \partial_n (L_2 \bullet f) dx_n \\ &= [L_2 \bullet f]_{-\infty}^{\infty} = 0 \quad \square \end{aligned}$$

¹Analytic continuation of generalized functions with respect to a parameter, Functional Analysis and Applications 6, 26-40

歴史

1. 佐藤幹夫: 代数解析の創始. 1960's — 1990's.
2. 柏原, 河合, J.Bernstein, Z.Mebkout, ... : D -加群の理論, 正則ホロノミック系 (regular holonomic systems), ... 代数幾何, 表現論への応用, ... 1970's — 現在.
3. D.Zeilberger, ... : 恒等式の証明および導出のホロノミック法. 1990's — 現在.
4. 大阿久, N.T, U.Walther, H.Tai, A.Leykin, 中村, 田島, 鍋島, ... : 計算 D -加群. 1990's – 現在.
5. 清, 竹村, 小山, N.T., 中山, 小原, 野呂, 橋口, 沼田, 栗木, 間野, 松原=許, ...² : ホロノミック勾配法 (holonomic gradient method (HGM)) およびホロノミック勾配降下法 holonomic gradient descent, 計算超幾何, 2010's — 現在.

²<https://www.math.kobe-u.ac.jp/OpenXM/Math/hgm/ref-hgm.html>

ホロノミック超関数

定義: 超関数 (distribution) f がホロノミックイデアルで零化される時 f を **ホロノミック超関数と呼ぶ**. f が通常関数であるときは, f を **ホロノミック関数と呼ぶ**. ホロノミック超関数の性質を厳密性をこだわらずに述べれば³,

1. ホロノミック超関数の一部の変数についての定積分 (パラメータ付き定積分) は残りの変数についてのホロノミック超関数となる.
2. ホロノミック超関数の和, 積 (定義できれば), 微分もまたホロノミック超関数.

例: $n = 1$, $x = x_1$. ヘビサイド関数

$Y(x) = 1$ ($x \geq 0$), $Y(x) = 0$ ($x < 0$) は $x\partial_x \bullet Y = x\delta(x) = 0$,
 $(x\partial_x - 1) \bullet xY(x) = 0$ を満たす

(機械学習によく使われる) ReLU (rectified linear unit) 関数

$\sigma(x) = xY(x)$ もホロノミック超関数. この定積分

$g(a, b, c) = \int_{\mathbb{R}^2} \exp(-au^2 - 2buv - cv^2)\sigma(u)\sigma(v)dudv$ は a, b, c についてのホロノミック関数.

³e.g.,

<https://www.math.kobe-u.ac.jp/HOME/taka/ascm2003-paper.pdf>

まとめ: ホロノミック超関数はどれ?

1. $\exp(f(x_1, \dots, x_n))$, ここで f は有理式.
2. $\sin(x)$.
3. $\frac{1}{\sin x}$ [ヒント] 定理: 常微分方程式
 $(a_m(x)\partial^m + \dots + a_0(x)) \bullet f = 0$, $a_i \in \mathbb{C}[x]$ の任意の解は特異
点集合 $\{x \mid a_m(x) = 0\}$ の外で正則.
4. $\frac{1}{1+\exp(-x)}$ (シグモイド関数)
5. $\exp(\exp(x))$ (指数関数の合成関数)
6. $\Gamma(x)$, [ヒント] $\Gamma(x)$ 関数は $x = -n$, $n \in \mathbb{N}_0$ に極を持つ.
7. x^a , ここで a は定数.
8. $|x|$,
9. $\int_{-\infty}^{+\infty} \exp(-xt^6 - t) dt$, $x > 0$.

2. 力学系としての最尤推定 (maximal likelihood estimation (MLE))

I は D_n のホロノミックイデアルで f を零化するとする. $R_n I$ は R_n の 0-次元イデアルとなる. $r := \dim_{K(x)} \frac{R_n}{R_n I}$ を **ホロノミックランク** と呼ぶ. $s_1, \dots, s_r : \frac{R_n}{R_n I}$ の基底. $F = (s_1 \bullet f, \dots, s_r \bullet f)^T$ とおくと次の **パッフィアン方程式**:

$$\frac{\partial F}{\partial x_i} = P_i(x) F \quad (2)$$

を満たす. ここで P_i は有理式係数の $r \times r$ 行列⁴.

ホロノミック勾配法 (holonomic gradient method (HGM)): ホロノミックな f に対して $Z(x') = \int_D f(x) dx_{m+1} \cdots dx_n$ を数値計算, $x' = (x_1, \dots, x_m)$.

(1) 積分イデアルを計算 (2) パッフィアン方程式を計算. (3a) F (とその微分達) をうまい点で評価. (3b) パッフィアン方程式を数値解析してその値を補外 (または補間).

⁴ P_i はグレブナー基底での簡約化で計算できる. コンパニオン行列.

最尤推定 (Maximal likelihood estimation (MLE))

例: 非正規化 Von-Mises 分布 ($S^1 \ni x$ 上の分布):

$$u(\theta, x) = \exp(\theta_1 \cos x + \theta_2 \sin x) = \exp(\theta_1 t_1 + \theta_2 t_2) \delta(t_1^2 + t_2^2 - 1).$$

正規化定数はホロノミック関数

$$Z(\theta) = \int_0^{2\pi} \exp(\theta_1 \cos x + \theta_2 \sin x) dx$$

$F = \left(Z, \frac{\partial Z}{\partial \theta_1} \right)^T$. パツフィアン方程式は

$$\begin{aligned} \frac{\partial F}{\partial \theta_1} &= \begin{pmatrix} 0 & 1 \\ \frac{\theta_1^2}{\theta_1^2 + \theta_2^2} & \frac{\theta_2^2 - \theta_1^2}{\theta_1(\theta_1^2 + \theta_2^2)} \end{pmatrix} F =: P_1 F \\ \frac{\partial F}{\partial \theta_2} &= \begin{pmatrix} 0 & \theta_2/\theta_1 \\ \frac{\theta_1 \theta_2}{\theta_1^2 + \theta_2^2} & \frac{-2\theta_2}{\theta_1^2 + \theta_2^2} \end{pmatrix} F =: P_2 F \end{aligned}$$

Fisher's MLE. X_i を観測データ. 尤度 (likelihood)

$\ell(\theta; X) = \prod_{i=1}^N \frac{u(\theta, X_i)}{Z(\theta)}$ を最大化する θ を見つけよ.

$f = \log \ell(\theta; X)$ を尤度の $\log$. 勾配降下法は $\theta = (\theta_1, \theta_2)$ を $(\text{new } \theta) = \theta + \alpha \nabla_{\theta} f$ により更新する. したがって⁵,

$$\dot{\theta} = \nabla_{\theta} f = \sum_{i=1}^N \frac{\nabla_{\theta} u(\theta, X_i)}{u(\theta, X_i)} - N \frac{\nabla_{\theta} F_1}{F_1}, \quad F_1 = Z$$

合成関数の微分公式とパッフィアン方程式により

$$\dot{F}_i = \dot{\theta}_1 (P_1 F)_i + \dot{\theta}_2 (P_2 F)_i$$

$$\dot{\theta}_1 = \sum_{i=1}^N \cos(X_i) - N \frac{(P_1 F)_1}{F_1} \quad (3)$$

$$\dot{\theta}_2 = \sum_{i=1}^N \sin(X_i) - N \frac{(P_2 F)_1}{F_1} \quad (4)$$

$$\dot{F}_i = \left(\sum_{i=1}^N \cos(X_i) - N \frac{(P_1 F)_1}{F_1} \right) (P_1 F)_i + \left(\sum_{i=1}^N \sin(X_i) - N \frac{(P_2 F)_1}{F_1} \right) (P_2 F)_i \quad (5)$$

$i = 1, 2.$

⁵U.Helmke, J.Moore, Optimization and Dynamical Systems, 1994

Von-Mises 分布の MLE

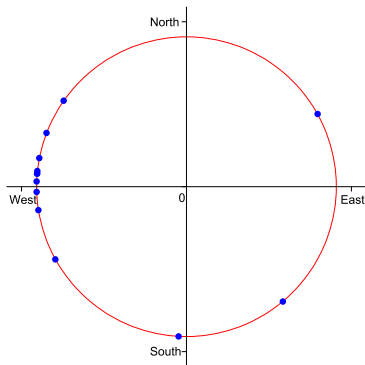


Figure: 札幌 10,000 メートル上空の風向, AM 9, 2011/1/1-2011/1/14
(1/11 欠損)

$\max_{\theta} \prod_{i=1}^{13} \frac{u(\theta, X_i)}{z(\theta)}$, ここで X_i は図の風向.

(3), (4), (5) のベクトル場

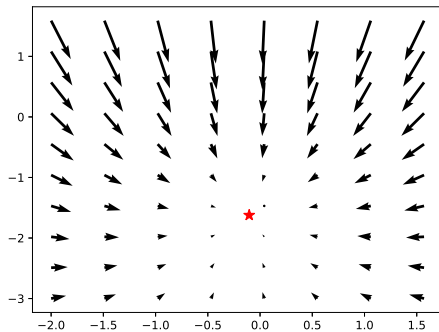


Figure: (θ_1, θ_2) 空間でのベクトル場

Von-Mises 分布の MLE

(3), (4), (5) を初期条件

$(\theta; F) = (-1.62, -0.1; 9.82246, -6.12855)$ で解くと,

$\theta = (\theta_1, \theta_2) = (-0.1038, -1.6228)$.

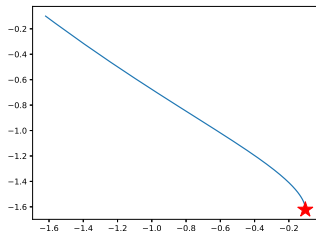
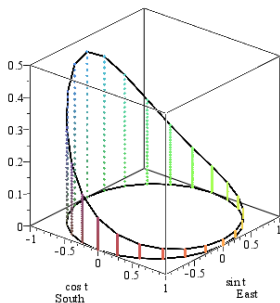


Figure: MLE により推定された風向分布

Figure: $\theta(t)$ の軌道

力学系としての MLE

定理 2. 非正規化分布 $u(\theta, x)$ がホロノミック関数^aで、正規化定数 $\int_{\Omega} u(\theta, x) dx$ が積分イデアルで零化されるなら、 $u(\theta, x)$ についての最尤推定問題は力学系として記述できる.

^a u がホロノミック超関数の場合条件の書き方がもっと煩雑に.

3. HGM の応用例

<https://www.math.kobe-u.ac.jp/OpenXM/Math/hgm/ref-hgm.html>

openxm hgm 検索.

1. 積分イデアル $(I + \partial_{m+1}D_n + \cdots + \partial_n D_n) \cap D_m$: を理論的に求める, または, より効率的なアルゴリズムで求める.
2. (巨大なサイズの) 常微分方程式の数値解析アルゴリズム⁶.

⁶<https://arxiv.org/abs/2111.10947> など

Neural Network (NN), Kernel method とは

$f(\theta, x) : \mathbb{R}^{d_0} \xrightarrow{\text{am}} \mathbb{R}^{d_1} \xrightarrow{\sigma} \mathbb{R}^{d_1} \xrightarrow{\text{am}} \mathbb{R}^{d_2} \xrightarrow{\sigma} \mathbb{R}^{d_2} \rightarrow \dots \xrightarrow{\sigma} \mathbb{R}^{d_{L+1}}$, “am” 達⁷ はパラメータ θ を持つアフィン写像. σ は activation(活性化関数).

$d_{L+1} = 1$, 最後の σ のみ id とする. Neural tangent kernel:

$$K(x, x') = \left\langle \frac{\partial f(x, \theta)}{\partial \theta}, \frac{\partial f(x', \theta)}{\partial \theta} \right\rangle \quad (6)$$

x_i 入力, y_i 出力, トレーニングデータ:

$\{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$.

kernel matrix: $H^* = (K(x_i, x_j)) \in \mathbb{R}^{N \times N}$. Kernel method

$$f(x) \sim (K(x, x_1), K(x, x_2), \dots, K(x, x_N))(H^*)^{-1}(y_1, y_2, \dots, y_N)^T. \quad (7)$$

⁷同じ記号を使うがすべて異なる

定理: Jacot et al 2018^a. NN の幅が $d_i \rightarrow \infty$ ならほとんどすべての θ について $K(x, x') \sim \Theta(x, x')$ となる.

^a<https://arxiv.org/abs/1806.07572>

参考文献⁸. Θ (NTK) の定義,

$$\Sigma^{(0)}(x, x') = x^T x', \quad (8)$$

$$\Lambda^{(h)}(x, x') = \begin{pmatrix} \Sigma^{(h-1)}(x, x) & \Sigma^{(h-1)}(x, x') \\ \Sigma^{(h-1)}(x', x) & \Sigma^{(h-1)}(x', x') \end{pmatrix} \quad (9)$$

$$\Sigma^{(h)}(x, x') = c_\sigma E_{(u,v) \sim N(0, \Lambda^{(h)})} [\sigma(u)\sigma(v)] \quad (10)$$

$$\dot{\Sigma}^{(h)}(x, x') = c_\sigma E_{(u,v) \sim N(0, \Lambda^{(h)})} [\dot{\sigma}(u)\dot{\sigma}(v)] \quad (11)$$

$$\Theta(x, x') = \Theta^{(L)}(x, x') = \sum_{h=1}^{L+1} \left(\Sigma^{(h-1)}(x, x') \prod_{h'=h}^{L+1} \dot{\Sigma}^{(h')}(x, x') \right) \quad (12)$$

⁸https://oumpy.github.io/blog/2020/04/neural_tangents.html,
Arora et al <https://arxiv.org/abs/1904.11955>

Dual activation

$E_{(u,v) \sim N(0, \Lambda^{(h)})}[\sigma(u)\sigma(v)]$ (dual activation of σ) は積分で書くと

$$\hat{E}[\sigma(u)\sigma(v)] = \int_{\mathbb{R}^2} \sigma(u)\sigma(v) \exp(x_{11}u^2 + 2x_{12}uv + x_{22}v^2) dudv$$

$$E_{(u,v) \sim N(0, \Lambda^{(h)})}[\sigma(u)\sigma(v)] = \hat{E}[\sigma(u)\sigma(v)] \frac{\sqrt{\det(x)}}{\pi}, \quad \Lambda^{(h)} = -\frac{1}{2}x^{-1}.$$

- ReLU $\sigma(u) = uY(u)$ ($Y(u)$ ヘビサイド関数).
- GeLU (Gaussian error linear unit): $\sigma(u) = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{u}{\sqrt{2}} \right) \right)$

Dual activation が計算できれば NTK が計算できる

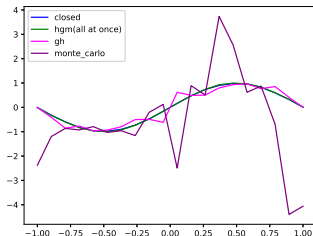
I.Han et al, Fast Neural Kernel Embeddings for General Activations, NeurIPS-2022.

A.Sakoda, N.Takayama, An Application of the Holonomic Gradient Method to the Neural Tangent Kernel, scml-2025⁹.

⁹<http://arxiv.org/abs/2410.23626>

迫田-N.T. の dual activation 計算法: 小山, 竹村による¹⁰ ガウス分布についての期待値の満たすホロノミック系の結果, および大阿久の制限アルゴリズム¹¹ を利用.

例: $\sigma(u) = \sin(u)$. $\sin(\pi x)$ の値を 15 点で与えたものを学習データとして NTK Θ で補間.



¹⁰<https://doi.org/10.1007/s13160-015-0166-8>, Calculation of Orthant Probabilities by the Holonomic Gradient Method (2015)

¹¹<https://doi.org/10.1006/aama.1997.0527> Algorithms for b -function, restrictions, and algebraic local cohomology groups (1997).

まとめ

1. ホロノミック関数, ホロノミック超関数はとてもよいクラスの関数.
2. 計算代数の手法を用いてこれらのパラメータ付き定積分の数値評価ができる (HGM).
3. ホロノミックな非正規化分布関数についての MLE は力学系として記述できる.
4. HGM はパラメータ付き積分が登場するいろんな分野に応用できる.