D-Modules in (Algebraic) Statistics

Anna-Laura Sattelberger (MPI-MiS Leipzig)

MCA 2021
Session: Differential operators in algebraic geometry and commutative algebra

July 16, 2021
Universidad de Buenos Aires
Maximum likelihood estimation via . . .

1. . . the holonomic gradient method
2. . . tropical geometry and Bernstein–Sato ideals
Holonomic Gradient Method
**Holonomic $D$-ideals**

- $D_n$ the **Weyl algebra** $\mathbb{C}[x_1, \ldots, x_n] \langle \partial_1, \ldots, \partial_n \rangle$
- $I$ a left $D$-ideal
- $M$ a left $D$-module

**Definitions**

- The **characteristic ideal** of a $D$-ideal $I$ is
  \[
  \text{in}_{(0,1)}(I) := \langle \text{in}_{(0,1)}(P) \mid P \in I \rangle \lhd \mathbb{C}[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n].
  \]
- $I$ is **holonomic** if $\dim(\text{in}_{(0,1)}(I)) = n$.
- $f \in M$ is **holonomic** if $\text{ann}_D(f) = \{ P \in D \mid P \bullet f = 0 \}$ is holonomic.
Holonomic functions

\[ M \in \text{Mod}(D_n) \quad \text{a space of functions} \]
\[ f \in M \quad \text{holonomic} \]

Facts & features

- encode by finite data (\(D_n\)-ideal + initial conditions)
  
  Example: sine is encoded by \((\partial^2 + 1) \bullet f = 0 \text{ and } (f(0) = 0, f'(0) = 1)\)

- closure properties (+, ·, \(\int\), *, \(\partial_i\), ...)

Examples

- algebraic/rational/hypergeometric/many special functions
- some probability distributions
- ... many more

Non-examples

Riemann zeta function \(\zeta\), \(\frac{1}{\sin}\), Lambert \(W\)-function
Holonomic gradient method

- numerical evaluation of holonomic functions
- keeping track of the gradient by the Pfaffian system

\[ \partial \bullet (f, f', \ldots, f^{(k-1)})^t = M \cdot (f, f', \ldots, f^{(k-1)})^t, \]

with \( M \in \text{Mat}_{k \times k}(\mathbb{C}(x)) \)

- several variables: Gröbner basis computations in the rational Weyl algebra
- holonomic gradient descent: minimization method based on the HGM
- freedom in choosing numerical methods
Maximum likelihood estimation

Input
Data \{x_1, \ldots, x_n\} + statistical model

Problem
Which parameters \( \theta \) of the model best explain the data, i.e., optimize the likelihood function \( \ell(\theta) := f(x_1 | \theta) \cdot \cdots \cdot f(x_n | \theta) \)?

Discrete case

- statistical experiment with \( N \) possible outcomes, probabilities \( p_1, \ldots, p_N \)
- data: \( (s_1, \ldots, s_N) \in \mathbb{N}^N \) count of outcome when repeating the experiment \( n = s_1 + \cdots + s_N \) many times
- maxima of \( \ell(p_1, \ldots, p_N) = \prod_{i=1}^{N} p_i^{s_i} \) : among the critical points of the log likelihood function \( \log \ell = \sum_{i=1}^{N} s_i \log p_i \)

\(^1\text{in the case of } i.i.d. \text{ random variables}\)
Sampling data from rotation groups

Fisher model

- family of probability distributions on SO(3) parametrized by $3 \times 3$-matrices $\Theta$
- for fixed $\Theta$, the density of the Fisher distribution equals

$$f_\Theta(Y) = \frac{1}{c(\Theta)} \cdot \exp(tr(\Theta^t \cdot Y)) \quad \text{for } Y \in \text{SO}(3)$$

- $c$ is the **normalizing constant**
- MLE for SO(3) via HGD in (Sei–Shibata–Takemura–Ohara–Takayama, 2013)

Other Lie groups than SO(3)

- $D$-ideal for SO($n$) studied in (Koyama, 2020)
- compact Lie groups in (Adamer–Lőrincz–S.–Sturmfels, 2020)
An example from medical imaging

The holonomic BFGS\(^2\) algorithm finds the MLE

\[ \hat{x}_1 = 20.072407, \quad \hat{x}_2 = 12.513841, \quad \hat{x}_3 = -6.510704. \]

\(^2\text{Broyden–Fletcher–Goldfarb–Shanno}\)
A Tropical and Bernstein–Sato Perspective
A geometric approach

Discrete statistical experiment: Flip a biased coin. If it shows *head*, flip again.

\[ x_0 \xrightarrow{s_0} x_1 \xrightarrow{s_1} x_0 \quad \text{and} \quad x_1 \xrightarrow{s_1} x_0 \xrightarrow{s_2} x_1 \]

\[ p_0 \quad \text{and} \quad p_1 \]

**Figure:** Staged tree modeling the experiment (Collazo–Görgen–Smith, 2018)

- \((s_0, s_1, s_2)\) count of outcome when repeating this experiment
- maximum likelihood estimate:

\[
\Psi(s_0, s_1, s_2) = \left( \frac{(2s_0 + s_1)^2}{(2s_0 + 2s_1 + s_2)^2}, \frac{(2s_0 + s_1)(s_1 + s_2)}{(2s_0 + 2s_1 + s_2)^2}, \frac{s_1 + s_2}{2s_0 + 2s_1 + s_2} \right)
\]

- parametrization of the model: \(\Delta_1 \to \Delta_2, (x_0, x_1) \mapsto (x_0^2, x_0x_1, x_1)\), where \(x_0, x_1 > 0, x_0 + x_1 = 1\)
- implicitization: \(M := V(p_0p_2 - (p_0 + p_1)p_1)\) smooth curve in \(\mathbb{P}^2\)
Bernstein–Sato ideals

\[ F = (f_1, \ldots, f_p) \in \mathbb{C}[x_1, \ldots, x_n]^p \quad \text{a tuple of polynomials} \]

Definition

The **Bernstein–Sato ideal** of \( F \) is the ideal \( B_F \) in \( \mathbb{C}[s_1, \ldots, s_p] \) consisting of polynomials \( b \) for which there exists \( P \in D_n[s_1, \ldots, s_p] \) such that

\[
P \star \left( f_1^{s_1+1} \cdots f_p^{s_p+1} \right) = b \cdot f_1^{s_1} \cdots f_p^{s_p}.
\]

Example: \( F = (x^2, x(1 - x), 1 - x) \)

Computed with the library dmod_lib in Singular:

\[
B_F = \left\langle \prod_{k=1}^{3} (2s_0 + s_1 + k) \cdot \prod_{\ell=1}^{2} (s_1 + s_2 + \ell) \right\rangle \triangleleft \mathbb{C}[s_0, s_1, s_2].
\]
MLE from a tropical and a Bernstein–Sato perspective

- rigorous explanation of the observed phenomenon in (S.–van der Veer, 2021)

Connecting three fields of research
- Bernstein–Sato theory
- likelihood geometry
- tropical geometry

Providing new tools for... 
- algebraic statistics
- high energy physics via scattering amplitudes (Sturmfels–Telen, 2020)
Revisiting the coin example

Implicit representation of the statistical model: smooth curve $\mathcal{M}$ in $\mathbb{P}^2$ defined by

$$f = \det \begin{pmatrix} p_0 & p_1 \\ p_0 + p_1 & p_2 \end{pmatrix} = p_0p_2 - (p_0 + p_1)p_1.$$

- $X \subseteq (\mathbb{C}^*)^2$ the very affine variety $\mathcal{M} \setminus \{p_0p_1p_2(p_0 + p_1 + p_2) = 0\}$
- rays in the tropical variety of $X$ are the rows of $^3$

$$\begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 1 \\ -2 & -2 & -1 \end{pmatrix}$$

- critical slopes: $V(2s_0 + s_1) \cup V(s_1 + s_2) \cup V(2s_0 + 2s_1 + s_2)$
- Bernstein–Sato ideal of the tuple $(x^2, x(1 - x), 1 - x)$ on $\mathbb{C}$:

$$\langle \prod_{k=1}^{3}(2s_0 + s_1 + k) \cdot \prod_{\ell=1}^{2}(s_1 + s_2 + \ell) \rangle \triangleleft \mathbb{C}[s_0, s_1, s_2]$$

$^3$computed with Gfan
$D$-modules are intriguing not only from a theoretical point of view; they provide useful techniques for concrete applications.
References


References II


Thank you very much for your attention!