D-Modules in (Algebraic) Statistics

Anna-Laura Sattelberger (MPI-MiS Leipzig)

SIAM AG 2021
MS92: Differential Equations in Algebraic Geometry and Beyond - Part II of III

August 20, 2021
Texas A&M University
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 $D_n$-ideal

$M$ a left $D_n$-module

**Definitions**

- The **characteristic ideal** of a $D_n$-ideal $I$ is

  $$\text{in}_{(0,1)}(I) := \langle \text{in}_{(0,1)}(P) \mid P \in I \rangle \triangleleft \mathbb{C}[x_1, \ldots, x_n, \partial_1, \ldots, \partial_n].$$

- $I$ is **holonomic** if $\dim(\text{in}_{(0,1)}(I)) = n$.

- $f \in M$ is **holonomic** if $\text{ann}_{D_n}(f) = \{P \in D_n \mid P \bullet f = 0\}$ 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 \((D_n\text{-ideal} + \text{initial conditions})\)
  
  **Example:** sine is encoded by \((\partial^2 + 1) \bullet f = 0\) and \((f(0) = 0, f'(0) = 1)\)

- closure properties \((+, \cdot, \int, \ast, \partial_i, \ldots)\)

Examples

- algebraic/rational/hypergeometric/many special functions
- some probability distributions
- \(\ldots\) many more

Non-examples

Riemann zeta function \(\zeta, \frac{1}{\sin},\) Lambert \(W\)-function
Evaluating and optimizing holonomic functions

Holonomic gradient method

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

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

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 \(\ell(\theta) := \prod_{i=1}^{n} f(x_i | \theta)\)?

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 \(\log \ell = \sum_{i=1}^{N} s_i \log p_i\)

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

Fisher model

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

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

This is the density with respect to the Haar measure $\mu$.
- The denominator is the **normalizing constant**. It is chosen such that

$$\int_{\text{SO}(3)} f_{\Theta}(Y) \mu(dY) = 1.$$ 

This requirement is equivalent to

$$c(\Theta) = \int_{\text{SO}(3)} \exp(\text{tr}(\Theta^t \cdot Y)) \mu(dY).$$

- MLE for $\text{SO}(3)$ via HGD in (Sei–Shibata–Takemura–Ohara–Takayama, 2013)
Generalizations

Other Lie groups than SO(3)

- $D_n^2$-ideal for SO($n$) studied in (Koyama, 2020)
- compact Lie groups in (Adamer–Lőrincz–S.–Sturmfels, 2020)

Theorem (Adamer–Lőrincz–S.–Sturmfels, 2020)

The annihilator of the normalizing constant is the Fourier–Laplace transform of a $D$-ideal that is obtained in terms of the equations of the group together with its Lie algebra data. Its associated $D$-module is simple holonomic.
An example from medical imaging

Figure: A dataset from a study in vectorcardiography (Downs–Liebman–Mackay, 1974)

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

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

\(^2\)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.

\[ \psi(s_0, s_1, s_2) = \begin{pmatrix} \frac{(2s_0 + s_1)^2}{(2s_0 + 2s_1 + s_2)^2}, & \frac{(2s_0 + s_1)(s_1 + s_2)}{(2s_0 + 2s_1 + s_2)^2}, & \frac{s_1 + s_2}{2s_0 + 2s_1 + s_2} \end{pmatrix} \]

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

\[\begin{align*}
M &:= V \left( p_0p_2 - (p_0 + p_1)p_1 \right) \text{ smooth curve in } \mathbb{P}^2
\end{align*}\]
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

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

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

Computed with the library `dmod_lib` in Singular:

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

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

- $X \subseteq (\mathbb{C}^*)^2$ the very affine variety $\mathcal{M} \setminus \{ p_0 p_1 p_2 (p_0 + p_1 + p_2) = 0 \}$
- 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 $\mathbb{C}$:

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

$^3$computed with Gfan
D-modules are intriguing not only from a theoretical point of view; they provide useful techniques for concrete applications.
<table>
<thead>
<tr>
<th>Reference</th>
<th>Author(s)</th>
<th>Title</th>
<th>Source</th>
</tr>
</thead>
</table>
References


Thank you very much for your attention!