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

MAX-PLANCK-GESELLSCHAFT

Outline

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

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

- I is **holonomic** if $\dim(\mathrm{in}_{(\mathbf{0},\mathbf{1})}(I)) = n$.
- $f \in M$ is **holonomic** if $\mathrm{ann}_{D_n}(f) = \{P \in D_n \mid P \bullet f = 0\}$ is holonomic.

Holonomic functions

$M \in \mathrm{Mod}(D_n)$	a space of functions
$f \in M$	holonomic

Facts & features

- encode by finite data (D_n-ideal + initial conditions)
  Example: sine is encoded by $(\partial^2 + 1) \bullet f = 0$ and $(f(0) = 0,\ f'(0) = 1)$
- closure properties $(+, \cdot, \int, *, \partial_i, \dots)$

Examples

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

Non-examples

Riemann zeta function ζ, $\frac{1}{\sin}$, Lambert W-function

Evaluating and optimizing holonomic functions

Holonomic gradient method
(Nakayama–Nishiyama–Noro–Ohara–Sei–Takayama–Takemura, 2011)

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

$$\partial \bullet \left(f, f', \ldots, f^{(k-1)}\right)^{\mathsf{t}} = M \cdot \left(f, f', \ldots, f^{(k-1)}\right)^{\mathsf{t}},$$

  with $M \in \mathrm{Mat}_{k\times k}(\mathbb{C}(x))$
- 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 θ of the model best explain the data, i.e., optimize the **likelihood function** $\ell(\theta) :=$ [1] $f(x_1 \mid \theta) \cdots f(x_n \mid \theta)$?

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 SO(3) parametrized by 3×3-matrices Θ
- For fixed Θ, the density of the Fisher distribution equals

$$f_{\Theta}(Y) = \frac{1}{c(\Theta)} \cdot \exp(\mathrm{tr}\left(\Theta^{\mathrm{t}} \cdot Y\right)) \qquad \text{for } \; Y \in \mathrm{SO}(3)$$

  This is the density with respect to the Haar measure μ.
- The denominator is the **normalizing constant**. It is chosen such that $\int_{\mathrm{SO}(3)} f_{\Theta}(Y)\mu(dY) = 1$. This requirement is equivalent to

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

- MLE for 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

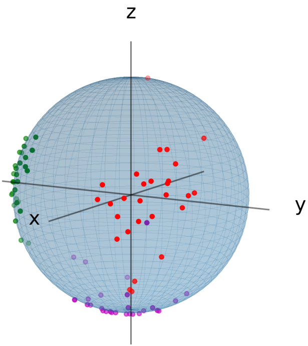

Figure: A dataset from a study in vectorcardiography (Downs–Liebman–Mackay, 1974)

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.

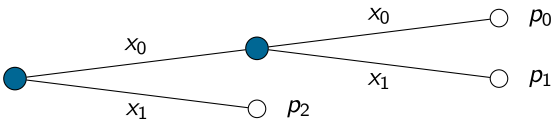

Figure: Staged tree modeling the experiment (Collazo–Görgen–Smith, 2018)

- (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: $\Delta_1 \to \Delta_2,\ (x_0, x_1) \mapsto (x_0^2, x_0 x_1, x_1)$, where $x_0, x_1 > 0,\ x_0 + x_1 = 1$
- implicitization: $\mathcal{M} := V(p_0 p_2 - (p_0 + p_1) p_1)$ smooth curve in $\mathbb{P}^2$

Bernstein–Sato ideals

$F = (f_1, \ldots, f_p) \in \mathbb{C}[x_1, \ldots, x_n]^p$ 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

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

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 $\mathbb{C}$:

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

[3] computed with `Gfan`

Conclusion

D-modules are intriguing not only from a theoretical point of view; they provide useful techniques for concrete applications.

References I

[ALSS20] Michael F. Adamer, Andras L. Lőrincz, Anna-Laura Sattelberger, and Bernd Sturmfels.
Algebraic Analysis of Rotation Data.
Algebraic Statistics, 11(2):189–211, 2020.

[Bud15] Nero Budur.
Bernstein–Sato ideals and local systems.
Ann. Inst. Fourier (Grenoble), 65(2):549–603, 2015.

[BvWZ21] Nero Budur, Robin van der Veer, Lei Wu, and Peng Zhou.
Zero loci of Bernstein–Sato ideals.
Invent. Math., 225:45–72, 2021.

[CGS18] Rodrigo A. Collazo, Christiane Görgen, and Jim Q. Smith.
Chain Event Graphs.
Computer Science & Data Analysis. Chapman & Hall/CRC, 1st edition, 2018.

[DGPS20] Wolfram Decker, Gert-Martin Greuel, Gerhard Pfister, and Hans Schönemann.
Singular 4-1-3—A computer algebra system for polynomial computations.
http://www.singular.uni-kl.de, 2020.

[DLM74] Thomas D. Downs, Jerome Liebman, and Wilma Mackay.
Statistical methods for vectorcardiogram orientations.
In *Vectorcardiography 2: Proc. XIth Int. Symp. Vectorcardiography*, pages 216–222, 1974.

[HNTT13] Hiroki Hashiguchi, Yasuhide Numata, Nobuki Takayama, and Akimichi Takemura.
The holonomic gradient method for the distribution function of the largest root of a Wishart matrix.
J. Multivariate Anal., 117:296–312, 2013.

[HS14] June Huh and Bernd Sturmfels.
Likelihood geometry.
In *Combinatorial algebraic geometry*, volume 2108 of *Lecture notes in mathematics*, pages 63–117. Springer, New York, 2014.

[Jen] Anders N. Jensen.
Gfan, a software system for Gröbner fans and tropical varieties.
Available at http://home.imf.au.dk/jensen/software/gfan/gfan.html.

References II

[Koy20] Tamio Koyama.
The annihilating ideal of the Fisher integral.
Kyushu J. Math, 74:415–427, 2020.

[LMM] Viktor Levandovskyy and Jorge Martín-Morales.
dmod_lib: A `Singular:Plural` library for algorithms for algebraic D-modules.
https://www.singular.uni-kl.de/Manual/4-2-0/sing_537.

[MS15] Diane Maclagan and Bernd Sturmfels.
Introduction to Tropical Geometry, volume 161 of *Graduate studies in mathematics*.
American Mathematical Society, Providence, R.I., 2015.

[NNN+11] Hiromasa Nakayama, Kenta Nishiyama, Masayuki Noro, Katsuyoshi Ohara, Tomonari Sei, Nobuki Takayama, and Akimichi Takemura.
Holonomic gradient descent and its application to the Fisher–Bingham integral.
Adv. in Appl. Math., 47(3):639–658, 2011.

[SST00] Mutsumi Saito, Bernd Sturmfels, and Nobuki Takayama.
Gröbner deformations of hypergeometric differential equations, volume 6 of *Algorithms and Computation in Mathematics*.
Springer-Verlag, Berlin, 2000.

[SST+13] Tomonari Sei, Hiroki Shibata, Akimichi Takemura, Katsuyoshi Ohara, and Nobuki Takayama.
Properties and applications of Fisher distribution on the rotation group.
J. Multivariate Anal., 116:440–455, 2013.

[ST20] Bernd Sturmfels and Simon Telen.
Likelihood Equations and Scattering Amplitudes.
arXiv:2012.05041 [math.AG], 2020.

[SvdV21] Anna-Laura Sattelberger and Robin van der Veer.
Maximum Likelihood Estimation from a Tropical and a Bernstein–Sato Perspective.
arXiv:2101.03570 [math.AG], 2021.

[Zei90] Doron Zeilberger.
A holonomic systems approach to special functions identities.
J. Comput. Appl. Math., 32(3):321–368, 1990.

Thank you very much for your attention!