Algebraic Optimization Degree

A Macaulay2 package

Marc Härkönen [1]

Benjamin Hollering [2]   Fatemeh Tarashi Kashani   Jose Israel Rodriguez [3]

[1]Georgia Tech

[2]North Carolina State University

[3]University of Wisconsin

Introduction

Introduction

Setting

Maximize/minimize an objective function Ψ over an algebraic variety X.

The algebraic degree of an optimization problem gives an algebraic measure of complexity of a problem

The algebraic degree of an optimization problem is an important invariant in applied algebraic geometry:

- nearest point problems (Draisma et al. 2016),
- maximum likelihood estimation (Catanese et al. 2006; Hoşten, Khetan, and Sturmfels 2005),
- semidefinite programming (Graf von Bothmer and Ranestad 2009).

A simple example

Let Y be a binomial random variable with 2 trials.

The probability mass function of Y is given by the map

$$p \mapsto ((1-p)^2, 2p(1-p), p^2) =: (p_0, p_1, p_2)$$

The Zariski closure of the image is the algebraic variety corresponding to the ideal

$$I = \langle 4p_0p_2 - p_1^2, 1 - p_0 - p_1 - p_2 \rangle \subseteq \mathbb{R}[p_0, p_1, p_2]$$

A simple example

Suppose we observe Y and collect the results in a vector (u_0, u_1, u_2).

The likelihood function is

$$\Psi(p_0, p_1, p_2) = p_0^{u_0} p_1^{u_1} p_2^{u_2}$$

Maximum likelihood estimation

Maximize Ψ subject to $(p_0, p_1, p_2) \in \mathbb{V}(I)$ and $0 \leq p_i \leq 1$ for $i \in \{1, 2, 3\}$.

Approach: find all complex critical points of Ψ. The number of critical points is the *maximum likelihood degree*.

General setting

- $X \subset \mathbb{C}^n$ an affine variety of codimension c
- $\Psi\colon X \to \mathbb{C}$ an objective function, with gradient $\nabla\Psi$

Definition

The critical ideal $\mathrm{Crit}_0(\Psi, X)$ is the ideal of the set of isolated critical points of Ψ on the regular locus of X. The degree of $\mathrm{Crit}_0(\Psi, X)$ is the optimization degree.

$$S := \left(\langle f_1, \ldots, f_N \rangle + \left\langle (c+1) \times (c+1) \text{ minors of } \begin{bmatrix} \nabla\Psi \\ \nabla f_1 \\ \vdots \\ \nabla f_N \end{bmatrix} \right\rangle \right) : I_{X_{\mathrm{sing}}}^{\infty},$$

where $f_1, \ldots, f_n$ generate the radical ideal of X. Let P denote the ideal of positive dimensional components of the variety of S. Then,

$$\mathrm{Crit}_0(\Psi, X) = S : P^{\infty}.$$

Euclidean distance degree

ED-degree

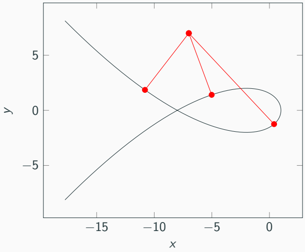
Problem

Given a variety $X \subseteq \mathbb{C}^n$ and a point $u \in \mathbb{C}^n$, find the point $x \in X$ that minimizes the Euclidean distance $\|x - u\|_2$.

The number of critical points of the Euclidean distance function for a *generic* point u is the Euclidean distance degree

Example

```
i1 : R = QQ[x,y];
i2 : I = ideal(27*y^2-(1-x)*(8+x)^2);
i3 : probabilisticEDDegree(I)
o3 = 5
i4 : symbolicEDDegree(I)
o4 = 5
```



ED-degree via projections

Let $X \subseteq \mathbb{P}^n$ have codimension ≥ 2. Let $\pi\colon \mathbb{P}^n \to \mathbb{P}^{n-1}$ be a rational map defined by a general linear map $\mathbb{C}^{n+1} \to \mathbb{C}^n$.

Theorem (Draisma et al. (2016))

The ED-degree of X is equal to the ED-degree of $\pi(X)$.

Example

```
i1 : R = ZZ/101[x_1..x_6]
i2 : M = genericSymmetricMatrix(R,3)
i3 : I = minors(2,M)
i4 : elapsedTime probabilisticEDDegree I
     -- 566.658 seconds elapsed
o4 = 13
i5 : elapsedTime projectionEDDegree I
     -- 0.115337 seconds elapsed
o5 = 13
```

Maximum likelihood degree

Introduction

- Probability simplex

$$\Delta_n := \{(p_0, \dots, p_n) : p_i \geq 0, \sum_i p_i = 1\}$$

- The statistical model is the solution set of homogeneous polynomials in $p_0, \dots, p_n$.
- Let V be the variety corresponding to the model.

Maximum Likelihood Estimation

Given a vector $u \in \mathbb{N}^{n+1}$ of observations, maximize

$$L = \frac{p_0^{u_0} + \cdots + p_n^{u_n}}{(p_0 + \cdots + p_n)^{u_0 + \cdots + u_n}}$$

subject to $(p_0, \dots, p_n) \in V \cap \Delta_n$.

Definition (Hoşten, Khetan, and Sturmfels (2005))

The ML-degree is the number of critical points of L for generic u.

Implicit form

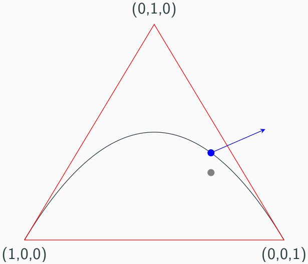
Example

Binomial random variable with 2 trials. Say we observe $u = (2, 5, 9)$.

```
i1 : R = QQ[p_0..p_2]
i2 : P = ideal(4*p_0*p_2-p_1^2)
i3 : MLequationsIdeal(P, {2,5,9})

o3 = ideal (23p  - 18p , 529p  - 81p )
               1      2       0      2

i4 : MLequationsDegree(P)
o4 = 1
```



Toric models

Definition

The scaled toric variety is the Zariski closure of the map $\phi_{A,c} : (\mathbb{C}^*)^d \to (\mathbb{C}^*)^r$ in $\mathbb{C}^r$ given by

$$\phi_{A,c}(\theta_1, \ldots, \theta_d) = (c_1\theta^{a_1}, \ldots, c_r\theta^{a_r}).$$

Example

```
i1 : A = matrix {{1,1,1,0,0,0,0,0,0}, {0,0,0,1,1,1,0,0,0},
{0,0,0,0,0,0,1,1,1},{1,0,0,1,0,0,1,0,0},{0,1,0,0,1,0,0,1,0}}
i2 : c = {1,2,3,1,1,1,1,1,1};
i3 : elapsedTime toricMLDegree(A, c)
     -- 0.00849565 seconds elapsed
o3 = 3
```

The corresponding computation using `MLequationsDegree` took 228 seconds.

Optimizations and future

Fritz John and Lagrange multipliers

- Let $A \in \mathbb{C}[x]^{m \times n}$ be a (generically) full-rank matrix, $m \geq n$.
- Computing $\langle n \times n$ minors of $A\rangle$ may be difficult.
- Lagrange/Fritz John:

$$\begin{bmatrix} z_1 & \cdots & z_n \end{bmatrix} A = 0$$

$$\sum_i c_i z_i = 1$$

Example

```
i1 : R = QQ[x_1..x_6];
i2 : M = genericSymmetricMatrix(R,3);
i3 : I = minors(2, M);
i4 : probabilisticEDDegree I
     -- 179.771 seconds elapsed
o4 = 13
i5 : probabilisticFritzJohnEDDegree I
     -- 4.95628 seconds elapsed
o5 = 13
```

Numerical algebraic geometry

Future: push beyond the boundaries of symbolic methods with numerical algebraic geometry

- Compute all critical points
- Find the optimal point
- Parametric homotopy
- `MonodromySolver`

Conclusion

Package features

- General optimization problem
  - Ideal
  - Largrange
- Euclidean Distance Degree
  - Probabilistic
  - Symbolic
  - Projective
    - Multidegree
    - Projections
    - Sections
  - Fritz John
- Maximum Likelihood degree
  - Symbolic
  - Parametric
  - Toric

Package available at
`https://github.com/Macaulay2/Workshop-2020-Cleveland/tree/ISSAC-AlgOpt/alg-stat/AlgebraicOptimization`

We thank the organizers, Federico Galetto, Courtney Gibbons, Hiram López, and Branden Stone, of the Macualay2 workshop at Cleveland State University for making this collaboration possible.

For more questions, join the link

https://bluejeans.com/550710311

on Jul 21, 18:00 EET

References

Catanese, Fabrizio et al. (2006). "The maximum likelihood degree". In: *Amer. J. Math.* 128.3, pp. 671–697. ISSN: 0002-9327. URL: http://muse.jhu.edu/journals/american_journal_of_mathematics/v128/128.3catanese.pdf.

Draisma, Jan et al. (2016). "The Euclidean distance degree of an algebraic variety". In: *Found. Comput. Math.* 16.1, pp. 99–149. ISSN: 1615-3375. DOI: 10.1007/s10208-014-9240-x. URL: https://doi.org/10.1007/s10208-014-9240-x.

Graf von Bothmer, Hans-Christian and Kristian Ranestad (2009). "A general formula for the algebraic degree in semidefinite programming". In: *Bull. Lond. Math. Soc.* 41.2, pp. 193–197. ISSN: 0024-6093. DOI: 10.1112/blms/bdn114. URL: https://doi-org.ezproxy.library.wisc.edu/10.1112/blms/bdn114.

Holme, Audun (1988). "The geometric and numerical properties of duality in projective algebraic geometry". In: *Manuscripta Math.* 61.2, pp. 145–162. ISSN: 0025-2611. DOI: 10.1007/BF01259325. URL: https://doi.org/10.1007/BF01259325.

Hoşten, Serkan, Amit Khetan, and Bernd Sturmfels (2005). "Solving the likelihood equations". In: *Found. Comput. Math.* 5.4, pp. 389–407. ISSN: 1615-3375. DOI: 10.1007/s10208-004-0156-8. URL: http://dx.doi.org/10.1007/s10208-004-0156-8.

Backup slides

ED-degree via multidegree

Theorem (Draisma et al. (2016))

The ED-degree of X is the sum of the polar classes of $\mathcal{N}$

Example

```
i1: R = QQ[x_0..x_3]
i2: J = ideal det(matrix{
  {x_0, x_1, x_2},
  {x_1, x_0, x_3},
  {x_2, x_3, x_0}});
i3: symbolicMultidegreeEDDegree J
     -- 0.263975 seconds elapsed
o3 = 13
i4: probabilisticMultidegreeEDDegree J
     -- 0.264061 seconds elapsed
o4 = 13
```

Projective varieties

The ED-degree of a projective variety $X \subseteq \mathbb{P}^n$ is the ED-degree of its affine cone in $\mathbb{C}^{n+1}$.

Definition

- The conormal variety $\mathcal{N}$ is the closure of the set

$$\{(x, u) \in \mathbb{P}^n \times (\mathbb{P}^n)^* : x \in X \setminus X_{\mathrm{sing}}, u \in T_x X\}$$

- We say X is in general position if $\mathcal{N}$ does not intersect the diagonal $\Delta \subseteq \mathbb{P}^n \times (\mathbb{P}^n)^*$.

Throughout the rest of the section, X is a projective variety in general position.

ED-degree via linear sections

Theorem

$$\mathrm{EDdegree}(X) = \begin{cases} \mathrm{EDdegree}(X \cap H), & \textit{if } \mathrm{codim}(X^*) \geq 2 \\ \mathrm{EDdegree}(X \cap H) + \deg(X^*), & \textit{if } \mathrm{codim}(X^*) = 1 \end{cases}$$

Notes

- Requires computing the projective dual X^*, i.e. the projection of $\mathcal{N}$ onto its second factor.
- Iterated application requires computation of $(X \cap H)^*$

Theorem (Holme (1988))

Let H be a hyperplane. If X^* has codimension ≥ 2, then

$$(X \cap H)^* = \pi_H(X^*),$$

where π_H is the projection with center point $H \in (\mathbb{P}^n)^*$

ML-degree of parametric models

In many applications, the statistical model is given as the image of a polynomial map

Example

```
i1 : R = QQ[p]
i2 : param = {(1-p)^2, 2*p*(1-p), p^2}
i3 : parametricMLIdeal(param, {2,5,9})
o3 = ideal(32p - 23)
i4 : parametricMLDegree(param)
o4 = 1
```