ALGEBRAIC OPTIMIZATION DEGREE

MARC HÄRKÖNEN, BENJAMIN HOLLERING, FATEMEH TARASHI KASHANI,
AND JOSE ISRAEL RODRIGUEZ

Abstract. The Macaulay2 [5] package `AlgebraicOptimization` implements methods for determining the algebraic degree of an optimization problem. We describe the structure of an algebraic optimization problem and explain how the methods in this package may be used to determine the respective degrees. Special features include determining Euclidean distance degrees and maximum likelihood degrees. To our knowledge, this is the first comprehensive software package combining different methods in algebraic optimization. The package is available at `https://github.com/Macaulay2/Workshop-2020-Cleveland/tree/ISSAC-AlgOpt/alg-stat/AlgebraicOptimization`.

1. Introduction

The algebraic degree of an optimization problem is an important invariant in applied algebraic geometry. It gives an algebraic measure of complexity to a problem and has been studied in the context of nearest point problems [3], maximum likelihood estimation [2, 6], and semidefinite programming [4].

The optimization degree [7] can be determined by computing the degree of an ideal. Let X denote an affine variety in $\mathbb{C}^n$. Given an objective function $\Psi : X \to \mathbb{C}$ with a gradient denoted by $\nabla\Psi$, our aim is to compute the ideal $\mathrm{Crit}_0(\Psi, X)$ of the set of isolated critical points of Ψ on the regular locus of X. We call the ideal $\mathrm{Crit}_0(\Psi, X)$ the (nondegenerate) *critical ideal* of X with respect to Ψ.

Suppose $(f_1, \dots, f_N)$ generate the radical ideal of X, and assume X has codimension c. Then $\mathrm{Crit}_0(\Psi, X)$ is given by these steps. First, consider the ideal S, which consists of the generators of the ideal of X along with minors of an augmented Jacobian matrix saturated by the ideal of the singular locus of X,

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

Then let P denote the ideal of positive dimensional components of the variety of S. Then,

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

In the special case when S is zero dimensional, this simplifies to $\mathrm{Crit}_0(\Psi, X) = S$. The (nondegenerate) *optimization degree* is the degree of $\mathrm{Crit}_0(\Psi, X)$. In this package we develop tools to generate these ideals and compute the respective degrees.

2. Package features

2.1. **Euclidean distance degree.**

A classic example of the optimization degree is the *Euclidean distance (ED) degree*. The ED degree of X is the optimization degree of the squared Euclidean distance function with a generic choice of *data* u. If X is a projective variety, we can use the following, which is more efficient than eq. (1) in some cases

$$\mathrm{Crit}_0(\Psi, X) = \left(I_X + \left\langle (c+2) \times (c+2) \text{ minors of } \begin{bmatrix} x \\ u \\ \mathrm{Jac}(I_X) \end{bmatrix} \right\rangle \right) : (I_{X_{\mathrm{sing}}}) \cdot I_Q)^{\infty}, \tag{2}$$

where $Q = \{x \in \mathbb{P}^{n-1} : x_1^2 + \cdots + x_n^2 = 0\}$ is the *isotropic quadric*, and $\mathrm{Jac}(I_X)$ is the Jacobian of the generators of I_X. For more details on this formulation see [3, Section 2]. Our package implements ED degree computations via eqs. (1) and (2) with the functions `probabilisticEDDegree`

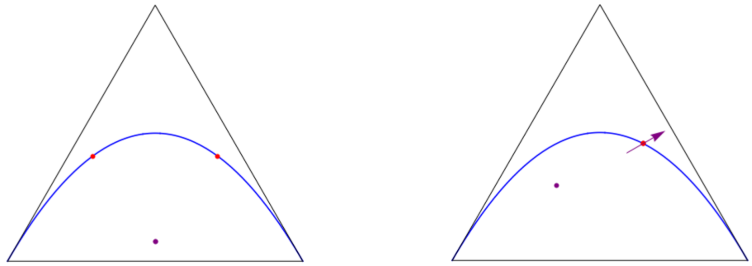

FIGURE 1. The variety X in Example 2 and related critical points given by $\mathrm{Crit}_0(\Psi, X)$. The left pictures a data point and the minimizers of Euclidean distance while the right pictures a data point and the corresponding MLE.

and `symbolicEDDegree`. The former chooses a random data point u, while the latter carries out computations symbolically. These are the main ED degree functions in our package.

If X is an irreducible projective variety in *general position*, there are several other ways to compute the ED degree: via multidegree, projections or sections. By [3, Thm. 5.4], the ED degree is the sum of the *multidegrees* of the conormal variety. The function `symbolicMultidegreeEDDegree` computes the multidegree symbolically, using the Hilbert polynomial, and `probabilisticMultidegreeEDDegree` computes the multidegree by counting points in random linear slices of the ambient space.

The function `projectionEDDegree` projects X into a smaller ambient space such that the projection has codimension 1. By [3, Cor. 6.1] the ED degree of the projected variety is equal to the ED degree of X. This can provide speedups when the codimension of X is large.

Example 1. Consider the Veronese surface in $\mathbb{P}^5$. We can confirm using different functions that the ED degree is thirteen.

```
i1 : R = QQ[x_0..x_5];
i2 : I = minors(2, matrix{{x_0, x_1, x_2}, {x_1,x_3,x_4}, {x_2,x_4,x_5}});
i3 : elapsedTime probabilisticEDDegree I
     -- 206.523 seconds elapsed
o3= 13
i4 : elapsedTime probabilisticMultidegreeEDDegree I
     -- 473.283 seconds elapsed
o4 = 13
i5 : elapsedTime projectionEDDegree I
     -- 5.99635 seconds elapsed
o5 = 13
```

$\diamond$

2.2. **Maximum likelihood degree.** The maximum likelihood estimate (MLE) of a statistical model is the minimizer of the *likelihood function* of the data. Tools of computational algebra can be used when the statistical model is an algebraic variety. The *maximum likelihood (ML) degree* is the number of complex critical points of the likelihood equations for generic data.

Example 2. Let $X \subseteq \mathbb{C}^3$ be defined by the ideal $I = \langle 4p_0p_2 - p_1^2 \rangle \subseteq \mathbb{C}[p_0, p_1, p_2]$. The intersection $X \cap \Delta_2$ is the set of all possible probability distributions of a binomial random variable Y with two trials. This is the blue curve pictured in Figure 1.

Suppose we observe Y and collect the results into a vector $u = (u_0, u_1, u_2)$ where u_i is the number of trials that resulted in i "heads". The likelihood function, i.e. the likelihood of observing u is

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

The critical points of this model are precisely the points in $\mathrm{Crit}_0(\Psi, X)$. In this case the ML degree is one, so there is a unique MLE. This can be confirmed by calling `MLequationsDegree I`. The data $u = (17, 29, 12)$ in purple and respective MLE in red are pictured on the right in Figure 1. The purple arrow shows the direction of the gradient of the likelihood function at the MLE. $\diamond$

In statistics, many models are given by a parametrization. Our package has methods to compute the ML degree for these models. More precisely, let $F : \mathbb{R}^d \to \mathbb{R}^{n+1}$ be a polynomial map whose image is a parametric model. Each coordinate f_i of F is a polynomial in the model parameters $\theta = (\theta_0, ..., \theta_d)$. Assuming the summation of f_i's is equal to one, the likelihood function is $f_0(\theta)^{u_0} f_1(\theta)^{u_1} \cdots f_n(\theta)^{u_n}$, where $u = (u_0, ..., u_n)$ is a vector of natural numbers. The function `parametricMLDegree` computes the ML degree for a parametric model. For more details on parametric likelihood equations see [6, Section 7].

Example 3. We can check that the ML degree of the twisted cubic model, given in parametric form, is 3.

```
i2 : R = QQ[t]; s=1;
i4 : F = {s^3*(-t^3-t^2-t+1), s^2*t, s*t^2, t^3};
i5 : parametricMLDegree (F)
o5 = 3
```

$\diamond$

2.3. **Toric models.** Toric models are a commonly used class of models in algebraic statistics which correspond to discrete exponential families in statistics. Well known examples include discrete graphical models and hierarchical models. Our package has specialized methods for these models which exploit their additional structure to compute their ML degree more quickly.

Toric models are typically given parametrically by a full rank matrix $A \in \mathbb{Z}^{d\times r}$ and a vector $c \in \mathbb{C}^r$. The *scaled toric variety*, denoted X_A^c, corresponding to the pair (A, c) is the Zariski closure of the map $\phi_{A,c} : (\mathbb{C}^*)^d \to (\mathbb{C}^*)^r$ in $\mathbb{C}^r$ given by

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

By Birch's Theorem, if the vector $(1, 1, \ldots 1) \in \mathrm{rowspan}(A)$ then the ML degree of X_A^c is the number of complex solutions to the equations

$$Au = nAp \text{ and } p \in X_A^c \setminus X(p_1p_2 \ldots p_r(p_1 + p_2 \ldots + p_r)) \tag{3}$$

for generic data vectors u [1, Prop. 7]. Our package computes this degree using the parametric description of the model with the function `toricMLDegree`. This method first chooses a random data point u and then forms the ideal, I_A^c of equations given by Equation 3 but with p replaced by $\phi_{A,c}(\theta)$. It then forms the critical ideal of the likelihood function Ψ in terms of the parameters by computing the saturation $\mathrm{Crit}_0(\Psi, X) = I_A^c : I(\theta_1\theta_2 \ldots \theta_d(c_1\theta^{a_1} + c_2\theta^{a_2} + c_r\theta^{a_r}))^\infty$.

Example 4. Let A and c be as they are below. The toric variety X_A^c is a scaled Segre embedding so the corresponding toric ideal is generated by 2×2 minors. The general method `MLequationsDegree` takes some time, whereas the specialized method `toricMLDegree` computes the ML degree quickly.

```
i2 : 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}}
i3 : c = {1,2,3,1,1,1,1,1,1};
i4 : R = QQ[p_1..p_9];
i5 : M = matrix {{p_1, p_4, p_7}, {p_2/2, p_5, p_8}, {p_3/3, p_6, p_9}}
i6 : I = minors(2, M);
i7 : elapsedTime MLequationsDegree I
     -- 228.037 seconds elapsed
```

```
o7 = 3
i8 : elapsedTime toricMLDegree(A, c)
     -- 0.00849565 seconds elapsed
o8 = 3
```

◇

2.4. Fritz John conditions and Lagrange multipliers. Instead of formulating the critical ideal using minors to specify a rank deficiency like in eq. (1), one can use a null vector method. We implement null vector methods involving *Lagrange multipliers* and *Fritz John conditions*.

Fritz John conditions are implemented for the ED degree computation. The determinental conditions in eqs. (1) and (2) are expressed by finding a (nonzero) kernel element of the augmented matrix $\begin{pmatrix} x-u \\ \mathrm{Jac}(I_W) \end{pmatrix}$, where I_W is an ideal generated by $\mathrm{codim}(I_W)$ polynomials, with I_X as a minimal prime. This is implemented in `symbolicFritzJohnEDDegree` and `probabilisticFritzJohnEDDegree`. These functions tend to work well when the number of generators of I is larger than the codimension. Corresponding functionality using Lagrange multipliers is implemented in `probabilistic-LagrangeMultiplierOptimizationDegree`.

```
i1 : R = QQ[x_1..x_6];
i2 : I = minors(2, matrix{{x_1, x_2, x_3}, {x_2, x_4, x_5}, {x_3, x_5, x_6}});
i3 : elapsedTime probabilisticEDDegree I
     -- 179.771 seconds elapsed
o3 = 13
i4 : elapsedTime probabilisticFritzJohnEDDegree I
     -- 4.95628 seconds elapsed
o4 = 13
```

Acknowledgements. We thank the organizers, Federico Galetto, Courtney Gibbons, Hiram López, and Branden Stone, of the Macualay2 workshop at Cleveland State University for initiating this collaboration. M. Härkönen is partially supported by the Vilho, Yrjö and Kalle Väisälä Foundation. B. Hollering was partially supported by the US National Science Foundation (DMS 1615660).

References

[1] C. Améndola, N. Bliss, I. Burke, C. R. Gibbons, M. Helmer, S. Hoşten, E. D. Nash, J. I. Rodriguez, and D. Smolkin. The maximum likelihood degree of toric varieties. *J. Symbolic Comput.*, 92:222–242, 2019.

[2] F. Catanese, S. Hoşten, A. Khetan, and B. Sturmfels. The maximum likelihood degree. *Amer. J. Math.*, 128(3):671–697, 2006.

[3] J. Draisma, E. Horobeţ, G. Ottaviani, B. Sturmfels, and R. R. Thomas. The Euclidean distance degree of an algebraic variety. *Found. Comput. Math.*, 16(1):99–149, 2016.

[4] H.-C. Graf von Bothmer and K. Ranestad. A general formula for the algebraic degree in semidefinite programming. *Bull. Lond. Math. Soc.*, 41(2):193–197, 2009.

[5] D. R. Grayson and M. E. Stillman. Macaulay2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/.

[6] S. Hoşten, A. Khetan, and B. Sturmfels. Solving the likelihood equations. *Found. Comput. Math.*, 5(4):389–407, 2005.

[7] P. Rostalski and B. Sturmfels. Dualities. In *Semidefinite optimization and convex algebraic geometry*, volume 13 of *MOS-SIAM Ser. Optim.*, pages 203–249. SIAM, Philadelphia, PA, 2013.