MULTI-DEGREES IN POLYNOMIAL OPTIMIZATION

KEMAL ROSE
MPI MIS, INSELSTRASSE 22, 04103, LEIPZIG, GERMANY

Abstract. We study structured optimization problems with polynomial objective function and polynomial equality constraints. The structure comes from a multi-grading on the polynomial ring in several variables. For fixed multi-degrees we determine the generic number of complex critical points. This serves as a measure for the algebraic complexity of the optimization problem. We also discuss computation and certification methods coming from numerical nonlinear algebra.

1. Introduction

Consider the following optimization problem:

$$\begin{aligned} &\text{minimize} \quad f_0(x),\ x \in \mathbb{R}^n \\ &\text{subject to } f_1(x) = \cdots = f_m(x) = 0 \end{aligned} \tag{1}$$

Here $f_0, \ldots, f_m$ are polynomials in the ring $\mathbb{R}[X]$ in n variables, partitioned into k subsets

$$X = \bigcup_{i=1}^{k} X_i,\ X_i = (x_{i,1}, \ldots, x_{i,n_i})$$

with $n = n_1 + \cdots + n_k$. To find the global optimizers of (1) we investigate the critical points of f_0 restricted to the feasible region. Since each global optimizer is amongst these critical points, it is an important problem in polynomial optimization to understand how they depend on the data $f_0, \ldots, f_m$. Following [NR09] we call the number of complex critical points of (1) the *algebraic degree of optimization*. These degrees constitute fundamental invariants of the variety $\{f_1 = \cdots = f_m = 0\}$, including the *euclidean distance* degree [DHO$^+$16], the *maximum likelihood degree* [SHK06], and are still subjects of active research [KKS21].

Example 1.1 $(n = 2,\ k = 1)$**.** The optimization problem (1) with

$$f_0 = x^2 + y^2,\ f_1 = -x^2 + y^2 + 1$$

has two real critical points displayed in figure 1.

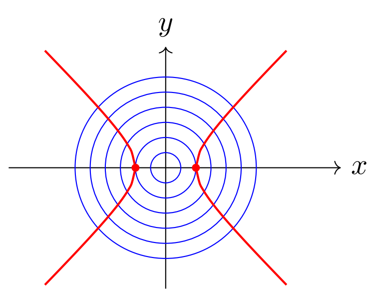

Figure 1. Level lines

In [NR09] the algebraic degree of optimization has been studied not for a specific choice of f_0, but instead it is determined for a generic choice of polynomials of fixed degrees. To weaken this restrictive genericity assumption it is desirable to replace the degree of a polynomial with a more refined notion. In this spirit we denote for every polynomial f in $\mathbb{R}[X]$ by $\deg(f) \in \mathbb{Z}^k_{\geq 0}$ its multi-degree. This is the vector of its degrees when considered as a polynomial in the k different sets of variables X_i.

The paper is structured as follows. In Section 2 we give homogeneous equations for the critical points. In Section 3 we determine the algebraic degree of optimization for a generic choice of polynomials $f_0, \dots, f_m$ with fixed multi-degrees, using methods from intersection theory. Our main result is Theorem 3.4. In the final section, Section 4, we discuss how numerical methods and bounds for the algebraic degree of optimization can be used in conjunction, in order to obtain optimizers of (1) together with a certificate for the correctness of the result. This uses an implementation of homotopy continuation [BT18], together with a certification method that we implemented, based on interval arithmetic [BRT20].

2. Critical point equations

From now on the polynomials f_i are generic polynomials with fixed, positive multidegrees $\deg(f_i) \in \mathbb{Z}^k_{\geq 0}$. A standard way of solving (1) is to apply Lagrange multipliers. In particular, the constrained optimization problem (1) is replaced with the unconstrained problem

$$\nabla \mathcal{L} = 0, \tag{2}$$

where $\nabla \mathcal{L}$ is the gradient of the Lagrangian

$$\mathcal{L} = f_0 - \lambda_1 f_1 - \cdots - \lambda_m f_m.$$

For the purposes of section 3 we will also consider the following alternative formulation. A smooth point x on the affine variety

$$V := \{x \in \mathbb{C}^n \mid f_1(x) = \cdots = f_m(x) = 0\}$$

is a critical point of f_0 if the rank of the Jacobian

$$M := \mathrm{Jac}(f_0, f_1, \dots, f_m) = (\nabla f_0, \nabla f_1, \dots, \nabla f_m)$$

drops at x. Note that, since all multi-degrees are strictly positive, by Bertini's theorem the variety V is smooth and of codimension m, so the Jacobian $(\nabla f_1, \dots, \nabla f_m)$ has full rank at every point in V. Let now W denote the affine variety determined by the vanishing of the maximal minors of M.

With the aim of computing the cardinality of the intersection $V \cap W$ using methods from intersection theory, we consider the closure $\mathcal{V}$ of V and $\mathcal{W}$ of W in the product of projective spaces

$$\mathcal{X} = \mathbb{P}^{n_1} \times \cdots \times \mathbb{P}^{n_k}.$$

We start by giving homogeneous defining equations: denote for every polynomial f in $\mathbb{R}[x]$ by $\tilde{f}$ its multihomogenization

$$\tilde{f} = x_{1,0}^{d_1} \cdots x_{k,0}^{d_k} f\left(\frac{X_1}{x_{1,0}}, \dots, \frac{X_k}{x_{k,0}}\right),$$

where the multi-degree of f is $(d_1, \dots, d_k)$. The vanishing locus $\mathcal{V}$ of $\tilde{f}_1, \dots, \tilde{f}_m$ is the Zariski-closure of V in $\mathcal{X}$. Analogously, let $\mathcal{M}$ denote the $n \times m+1$ matrix

$$\mathcal{M} = \begin{pmatrix} \frac{\partial}{\partial x_{1,1}}\tilde{f}_0 & \cdots & \frac{\partial}{\partial x_{1,1}}\tilde{f}_m \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{1,n_1}}\tilde{f}_0 & \cdots & \frac{\partial}{\partial x_{1,n_1}}\tilde{f}_m \\ \vdots & \vdots & \vdots \\ \frac{\partial}{\partial x_{k,1}}\tilde{f}_0 & \cdots & \frac{\partial}{\partial x_{k,1}}\tilde{f}_m \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{k,n_k}}\tilde{f}_0 & \cdots & \frac{\partial}{\partial x_{k,n_k}}\tilde{f}_m \end{pmatrix}.$$

The $m+1$ by $m+1$ minors of $\mathcal{M}$ are the homogenizations of the minors of M. Their vanishing locus is the closure $\mathcal{W}$ of W in $\mathcal{X}$.

3. Counting critical points

Our first observation is the following:

Proposition 3.1. *The intersection $\mathcal{V} \cap \mathcal{W}$ is finite.*

Proof. Assume that there is a curve C contained in $\mathcal{V} \cap \mathcal{W}$. Then C intersects the vanishing locus of $\tilde{f}_0$. Equivalently there is a point x contained in $\mathcal{V} \cap \{\tilde{f}_0 = 0\}$ where the maximal minors of $\mathcal{M}$ vanish. Since by Bertini's Theorem the intersection $\mathcal{V} \cap \{\tilde{f}_0 = 0\}$ is a smooth variety, the maximal minors of the matrix

$$\operatorname{Jac}(\tilde{f}_0, \ldots, \tilde{f}_m) = \begin{pmatrix} \frac{\partial}{\partial x_{1,0}} \tilde{f}_0 & \cdots & \frac{\partial}{\partial x_{1,0}} \tilde{f}_m \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{k,n_k}} \tilde{f}_0 & \cdots & \frac{\partial}{\partial x_{k,n_k}} \tilde{f}_m \end{pmatrix}$$

do not vanish simultaneously. This matrix differs from the matrix $\mathcal{W}$ by some additional rows corresponding to the differentials $\frac{\partial}{\partial x_{1,0}}, \ldots, \frac{\partial}{\partial x_{k,0}}$. However, as a consequence of the Euler relation, the rank of $\mathcal{M}$ and $\operatorname{Jac}(\tilde{f}_0, \ldots, \tilde{f}_m)$ coincide on $\mathcal{V}$, away from $\cup_{i=1}^{k} \{x_{i,0} = 0\}$. This gives us the desired contradiction. $\square$

Applying a Bertini-type argument to (2) shows that the intersection $\mathcal{V} \cap \mathcal{W}$ is transversal, implying the following lemma:

Lemma 3.2. *f_0 has finitely many critical points on V. Their number is identified with the intersection product $[\mathcal{V}][\mathcal{W}]$ in the cohomology ring $A^*(\mathcal{X}) = \mathbb{Z}[Y_1, \ldots, Y_k]/(Y_1^{n_1+1}, \ldots, Y_k^{n_k+1})$.*

As $\mathcal{V}$ is a transversal intersection of generic global sections of the line bundles $\mathcal{O}(\deg(f_j))$ on $\mathcal{X}$, its class in $A^*(\mathcal{X})$ is the product of Chern classes

$$[\mathcal{V}] = \prod_{j=1}^{m} c_1(\mathcal{O}(\deg(f_j))).$$

We now compute the class $[\mathcal{W}]$. The matrix $\mathcal{M}$ may be considered as a collection of linear maps parametrized by $\mathcal{X}$. In fact, each block

$$\begin{pmatrix} \frac{\partial}{\partial x_{i,1}} \tilde{f}_0 & \cdots & \frac{\partial}{\partial x_{i,1}} \tilde{f}_m \\ \vdots & \ddots & \vdots \\ \frac{\partial}{\partial x_{i,n}} \tilde{f}_0 & \cdots & \frac{\partial}{\partial x_{i,n_1}} \tilde{f}_m \end{pmatrix}$$

of $\mathcal{M}$ has elements of multi-degree $\deg(f_j) - e_i$ in the j-th cloumn. e_i denotes the i-th vector of unity. In particular it determines a map of vector bundles

$$\mathcal{O}(e_i)^{\oplus n_i} \longrightarrow \bigoplus_{j=0}^{m} \mathcal{O}(\deg(f_j))$$

on $\mathcal{X}$. Here $\mathcal{O}(e_i)$ is the $i-$th canonical line bundle on the product of projective spaces $\mathcal{X}$. We transpose $\mathcal{M}$ to get a map of bundles

$$\bigoplus_{j=0}^{m} \mathcal{O}(-\deg(f_j)) \longrightarrow \bigoplus_{i=1}^{k} \mathcal{O}(-e_i)^{\oplus n_i}.$$

On a smooth variety, and under the condition that the codimension of the degeneracy locus $\mathcal{W}$ of a morphism $\mathcal{M}$ of vector bundles is appropriate, the Thom-Porteous-Giambelli formula [Eis16] applies to determine the classes $[\mathcal{W}]$ in $A^*(\mathcal{X})$. For our choice of $\mathcal{M}$ the formula identifies $[\mathcal{W}]$ with the $n-m$-graded part of the quotient

$$\frac{c(\bigoplus_{i=1}^{k}\mathcal{O}(-e_i)^{\oplus n_i})}{c(\bigoplus_{j=0}^{m}\mathcal{O}(-\deg(f_j)))}$$

of the total Chern classes of $\bigoplus_{i=1}^{k}\mathcal{O}(-e_i)^{\oplus n_i}$ and $\bigoplus_{j=0}^{m}\mathcal{O}(-\deg(f_j))$ respectively. Let Y_i denote the $i-$th hyperplane class in $\mathcal{X}$. Then $c(\mathcal{O}(e_i)) = 1 + Y_i$ and we obtain

Lemma 3.3. *The class $[\mathcal{W}]$ in $A^*(\mathcal{X})$ is the degree $n-m$ homogeneous part of the expression*

$$\Sigma := \frac{\prod_{i=1}^{k}(1-Y_i)^{n_i}}{\prod_{j=0}^{m} 1-\sum_{i=1}^{k}\deg(f_j)_i Y_i.}$$

By abuse of notation we view Σ as a power series in the formal variables $Y_1,\dots,Y_k$. Σ is a sum of homogeneous polynomials in $Y_1,\dots,Y_k$:

$$\Sigma = \sum_{i=0}^{\infty}\Sigma_i,\ \Sigma_i \in \mathbb{Z}[Y_1,\dots,Y_k]_{(i)}.$$

The two Lemmas above allow us to compute the algebraic degree of optimization:

Theorem 3.4. *The number of critical points of f_0 on the variety $V = V(f_1,\dots,f_m)$ is the coefficient of the monomial $Y_1^{n_1}\cdots Y_k^{n_k}$ in the product*

$$\Sigma_{n-m}\prod_{j=1}^{m}\left(1-\sum_{i=1}^{k}\deg(f_j)_i Y_i\right).$$

Example 3.5. Consider the problem of optimizing a quadratic function f_0 over two quartic constraints f_1, f_2 in 4 variables x_1, x_2, y_1, y_2. For a generic choice of such functions the number of critical points is 544 [NR09, Theorem 2.2]. Compare this to the the case where f_0 is bilinear and both equality constraints are biquadratic. To apply theorem 3.4 let X and Y denote two formal variables. First we expand the power series

$$\frac{(1-X)^2(1-Y)^2}{(1-(X+Y))(1-2(X+Y))^2}$$

and compute the degree 2 homogeneous part $8X^2 + 18XY + 8Y^2$. The algebraic degree of optimization is the coefficient 208 of the monomial X^2Y^2 in $(2X+2Y)^2(8X+18XY+8Y)$. As we will verify below, the bound 208 is sharp.

4. Numerical nonlinear algebra

In the following we demonstrate how bounds on the algebraic degree of optimization can be used in conjunction with numerical methods, in order to compute optimizers for (1) together with a certificate for the correctness of the result. The basic idea is to solve the Lagrangian system (2) using numerical methods such as the software package `HomotopyContinuation.jl`. Although in principle it is possible to find all solutions, since this method involves numerical computations we might fail to find all critical points. There is however a certification method implemented. This feature gives *lower bounds* on the number of obtained critical points.

Example 4.1. Coming back to example 3.5 we consider an optimization problem with objective function of bidegree $(1,1)$ and equality constraints of bidegree $(2,2)$ in the varible groups x_1, x_2 and y_1, y_2:

$f_0 = 4y_2x_1 + 17y_2x_2 + 13x_1y_1 + 48x_2y_1 - 15y_2 + 6x_1 + 23x_2 + 38y_1 + 24$
$f_1 = x_1^2y_1^2 + x_1^2y_2^2 + x_2^2y_1^2 + x_2^2y_2^2 + 2x_1^2 + 3x_2^2 + 5y_1^2 + 7y_2^2 - 100,$
$f_2 = 20y_2^2x_1^2 - 18y_2^2x_1x_2 - 5y_2^2x_2^2 + 23y_2x_1^2y_1 + 28y_2x_1x_2y_+ 49y_2x_2^2y_1 - 8x_1^2y_1^2 + 46x_1x_2y_1^2 + 18x_2^2y_1^2 +$

$30y_2^2x_1 + 26y_2^2x_2 - 19y_2x_1^2 + 19y_2x_1x_2 - 9y_2x_1y_1 + 38y_2x_2^2 + 25y_2x_2y_1 - 17x_1^2y_1 + 28x_1x_2y_1 - 2x_1y_1^2 - 20x_2^2y_1 + 23x_2y_1^2 + 35y_2^2 + 46y_2x_1 + 29y_2x_2 + 4y_2y_1 + 6x_1^2 + 42x_1x_2 + 32x_1y_1 - 12x_2^2 - 13x_2y_1 + 39y_1^2 + 31y_2 + 50x_1 - 7x_2 + 22y_1 - 3$.

The optimization problem has exactly 208 complex critical points, 22 of which are real.

The following Julia code solves the Lagrangian formulation (2) of the optimization problem using the software package `HomotopyContinuation.jl`. It finds 208 critical points, certifies that all of them are distinct and that exactly 22 of them are real. This shows that the upper bound given by Theorem 3.4 is attained. In particular, by comparing the values of f_0 on the real ones, the optimizer can be determined.

```
@var x_1 x_2 y_1 y_2 l1 l2;
variables = [x_1, x_2, y_1, y_2, l1, l2];
f_0, f_1, f_2 = ...;
L = f0 - l1 * f1 - l2 * f2;
∇L = differentiate(L, variables);
result = solve(System(∇L));
certificate = certify(∇L, result);
reals = real_solutions(result);
obj_values =
map(s -> evaluate(f0, [x_1, x_2, y_1, y_2]
=> s[1:4]), reals);
minval, minindex = findmin(obj_values);
minarg = reals[minindex][1:4];
```

The objective function attains its minimal value -234.876 at the point $(0.011, 0.019, -6.355, 0.774)$.

Outlook. Theorem 3.4 can be stated analogously when taking the closure of V and W not in a product of projective spaces, but any toric variety. It is desirable to obtain a version of the theorem that holds for polynomials that are generic relative to their Newton polytopes or even generic relative to their supports. In particular, it is a natural question to ask when the BKK bound of the Lagrangian formulation (2) coincides with the algebraic degree of polynomial optimization. This is the subject of our next paper.

From a computational perspective it would be interesting to make the main theorem 3.4 effective by finding, for prescribed multidegrees, an instance of the problem (1) for which all critical points can be easily computed. This would provide an effective way to solve (1) by the means of homotopy continuation.

References

[BRT20] Paul Breiding, Kemal Rose, and Sascha Timme. Certifying zeros of polynomial systems using interval arithmetic, 2020.

[BT18] Paul Breiding and Sascha Timme. HomotopyContinuation.jl: A Package for Homotopy Continuation in Julia. In *Mathematical Software – ICMS 2018*, pages 458–465, Cham, 2018. Springer International Publishing.

[DHO+16] Jan Draisma, Emil Horobeţ, Giorgio Ottaviani, Bernd Sturmfels, and Rekha R. Thomas. The Euclidean Distance Degree of an Algebraic Variety. *Foundations of Computational Mathematics*, 16(1):99–149, 2016.

[Eis16] Harris Eisenbud. *3264 and All That – a Second Course in Algebraic Geometry*. Cambridge University Press, London, 2016.

[KKS21] Kaie Kubjas, Olga Kuznetsova, and Luca Sodomaco. Algebraic degree of optimization over a variety with an application to p-norm distance degree, 2021.

[NR09] Jiawang Nie and Kristian Ranestad. Algebraic degree of polynomial optimization. *SIAM Journal on Optimization*, 20:485—502, 2009.

[SHK06] Bernd Sturmfels, Serkan Hoşten, and Amit Khetan. The Maximum Likelihood Degree. *American Journal of Mathematics*, 128(3):671–697, 2006.

Email address: KEMAL.ROSE@MIS.MPG.DE