21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

ON THE ROLE OF THE CRITICAL VALUE POLYNOMIAL IN ALGEBRAIC OPTIMIZATION.

BERNARD HANZON AND ANDREI MUSTATA*

1. Introduction. In many applications of mathematics, including areas such as financial mathematics and control theory, optimization plays an important role. Often the criterion function is a polynomial or a rational function of the unknowns, which we will take to be real numbers here. In such a case the first order conditions for optimality can be written in the form of a polynomial system of equations and we speak of an algebraic optimization problem. The solutions of the polynomial system are called the critical points of the criterion function. The corresponding values of the criterion function are called critical values. There are only a finite number of such critical values and one can construct a univariate nonzero polynomial which is zero on the critical value set. Such a polynomial will be called a critical value polynomial (CVP).

In the talk we will explain methods to obtain such a critical value polynomial and some generalizations and their usage for determining the (global!) optimal value of the criterion.

This talk is partly based on the Masters project work of Marketa Adamova.

2. Optimization of a polynomial. Let us start with an example. Consider the univariate polynomial $f(x) = x^4 - 3x^2 + 2$. Its first derivative is $f'(x) = 4x^3 - 6x$ and its critical points, i.e. the zeros of the derivative, are $x = 0$ and $x = +/-\sqrt{\frac{3}{2}}$. The *critical values* are the corresponding function values: $f(0) = 2; f(+/-\sqrt{\frac{3}{2}}) = -\frac{1}{4}$. Now consider the auxiliary variable ϕ and consider the system of two equations given by

$$f'(x) = 0; \phi = f(x).$$

We can derive a relation between the powers of ϕ using the equation $f'(x) = 0 \leftrightarrow x^3 = \frac{3}{2}x$. It follows that $\phi = x^4 - 3x^2 + 2 = -\frac{3}{2}x^2 + 2$ and similarly $\phi^2 = -\frac{21}{8}x^2 + 4$. Using this we obtain $\phi^2 - \frac{7}{4}\phi - \frac{1}{2} = 0 \leftrightarrow (\phi + \frac{1}{4})(\phi - 2) = 0$. Such a polynomial will be called a *critical value polynomial (CVP)* as all the critical values are zeros of this polynomial. Note that a critical value polynomial can be computed *without* calculating the critical points! The same holds for multivariate polynomials. In fact we have:

THEOREM 2.1 (Algebraic Sard Theorem). *Let $f(x_1, x_2, \ldots, x_n)$ be a (possibly complex) polynomial. The number of critical values (including those attained at complex critical points) is finite. The ideal $< \frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}, \phi - f >$ contains a univariate polynomial in the variable ϕ. Such a polynomial is a critical value polynomial.*

Note that the theorem holds even if the number of critical points is infinite! Consider for example the bivariate polynomial $f(x_1, x_2) = (x_1x_2 - 1)^2$. Then the set of critical points consists of the union of the origin with the hyperbola given by the equation $x_1x_2 = 1$. This is clearly an infinite set. However taking $\phi = f(x_1, x_2)$ and using $x_1x_2^2 = x_2; x_1^2x_2 = x_1$ we obtain $\phi^2 - \phi$ as critical value polynomial and the critical values are 0 and 1.

*Dept. of Mathematics, University College Cork, e-mail for correspondence b.hanzon@ucc.ie
Supported by Science Foundation Ireland under grant numbers 07/MI/008 and RFP2007-MATF802

ISBN: 978-90-367-6321-9

3. Rational function optimization. Consider a rational function $f(x) = \frac{p(x)}{q(x)}$, $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, where p, q are polynomials with real coefficients. Suppose we want to find the *infimum* of f. Note that the infimum does *not* have to be attained even if it is finite. This phenomenon already occurs for bivariate polynomials. Consider for instance $f(x_1, x_2) = p(x_1, x_2) = x_1^2 + (x_1x_2 - 1)^2$. Clearly the infimum is zero but this value is not attained as it would require $x_1 = 0$, but $f(0, x_2) = 1 \neq 0$.

We will now present a redefinition of the problem which has a number of theoretical advantages both for rational functions as well as for general polynomial functions. We will make some assumptions:

- For the moment we will assume that the infimum is finite. We do have a way to test this once we have developed our approach for the finite infimum case.
- Without loss of generality we will assume that polynomials p, q do not have a common factor.
- Under these conditions q will not change sign (as was shown in [2]). We will assume without loss of generality that q is non-negative.

Under these assumptions, we have

$$\inf(f) = \max\{\gamma | \forall x \in \mathbb{R}^n : p(x) - \gamma.q(x) \geq 0\}$$

In order to handle "points at infinity" in the same way as ordinary points, we homogenize the polynomials involved: let $2d \geq \deg(p), \deg(q)$, $d \in \mathbb{N}$ and define $P(x_0, x_1, \ldots, x_n) = x_0^{2d} p(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_0})$ and similary $Q(x_0, x_1, \ldots, x_n) = x_0^{2d} q(\frac{x_1}{x_0}, \ldots, \frac{x_n}{x_0})$. Then P, Q are homogeneous of degree $2d$. Because $2d$ is even we obtain

$$\inf(f) = \max\{\gamma | \forall x \in \mathbb{R}^{n+1} : P(x) - \gamma.Q(x) \geq 0\}$$

Because of the homogeneity we can actually restrict the points x to a so-called Minkowski ball $S_{2\alpha} = \{x = (x_0, x_1, \ldots, x_n) | x_0^{2\alpha} + x_1^{2\alpha} + \ldots + x_n^{2\alpha} = 1\}$ where $\alpha \in \mathbb{N}$. Using the Lagrangian approach the first order conditions that are obtained are: $\frac{\partial(P - \gamma Q)}{\partial x_i} - 2\alpha\lambda x_i^{2\alpha - 1} = 0$, $i = 0, 1, 2, \ldots, n$. In case 2α is taken to be larger than the degree of $P - \gamma Q$ as a multivariate polynomial in x, this set of equations actually forms a Groebner basis for any fixed $\lambda \neq 0$. See for instance [1].

Now constructive algebra techniques can be used (such as Groebner basis techniques or resultant methods) to eliminate the variables $x_0, x_1, \ldots, x_n$ from these equations. A polynomial $F(\gamma, \lambda)$ will be obtained. We have:

THEOREM 3.1. *Let $m(\gamma)$ denote the coefficient of the smallest power of λ in $F(\gamma, \lambda)$, when viewed as a polynomial in λ. Then $m(\gamma)$ is a polynomial which has the (finite) infimum of f as one of its zeros.* The polynomial $m(\gamma)$ will be called a generalized critical value polynomial and its zeros will be called generalized critical values of the optimization problem.

4. Identifying the infimum among the generalized critical values. Once the generalized critical value polynomial is determined, one can compute its real zeros. There are finitely many of them and the infimum of the function is among them. However we still need to determine which one is the actual infimum. That the smallest real critical value of a function can be *smaller* than the infimum of the function can be seen from a simple example: Let $f(x) = (x^2 + 1)^2$ be a univariate polynomial. Clearly the minimum is attained at $x = 0$ and is equal to $f(0) = 1$. However the critical value polynomial obtained by eliminating the variable x from $f'(x) = 0; \phi = f(x)$, is equal to $\phi^2 - \phi$ and $\phi = 0$ and $\phi = 1$ are the critical values. So the smallest real critical value is zero and smaller than the infimum of the function.(Of course the critical points corresponding to the critical value zero are the complex points $x = +/-i$!!)

The question arises as to how we can determine which generalized critical value is the infimum value of the function. As the number of candidates is finite, all that is needed is sufficiently tight upper and lower bounds on the infimum!

5. Upper bound determination. First of all any reasonable heuristic method to find the infimum of f will attempt to construct points at which the function value of f is small. Therefore any such method will give an upper bound on the infimum. Whether the resulting upper bound is sufficiently tight cannot be guaranteed in advance in general. One such method is a Monte Carlo or Quasi Monte Carlo method in which the function is minimized over a finite grid of points. The idea here would be that by increasing the sample size of the MC or QMC eventually a point will be found in which the function value is sufficiently close to the infimum to rule out all generalized critical values larger than the infimum.

Remark. Alternatively it is possible to use the fact that for any value γ larger than the infimum, the homogeneous polynomial $P(x) - \gamma.Q(x)$ will have negative values; this then also holds if the domain of that function is restricted to the Minkowski ball $S_{2\alpha}$. On that ball $P - \gamma.Q$ will satisfy a Lipschitz condition which can be used in combination with the compactness of the ball to determine whether the minimum on the ball is negative or not. This can then be used to construct a finite-step decision procedure to determine which generalized critical value is the infimum. However this decision procedure may be slow and one may only want to use it if other methods do not give a conclusive answer. This needs to be investigated further.

6. Lower bound determination. Lower bound determination can be approached in several different ways. Here we propose to use a classical result from Polya:

THEOREM 6.1 (Polya). *Consider the $n-$dimensional simplex $\Delta_{n+1} = \{x \in \mathbb{R}^{n+1}_{\geq 0} \mid \sum_{i=0}^{n} x_i = 1\}$. If a homogeneous polynomial $R(x_0, x_1, \ldots, x_n)$ is (strictly!) positive on the simplex Δ_{n+1} then there exists a positive integer N such that $R(x_0, \ldots, x_n).(x_0 + x_1 + \ldots + x_n)^N$ is a (homogeneous) polynomial with all coefficients non-negative.*

Remarks

- Clearly a polynomial with all nonzero coefficients positive, is nonnegative on Δ_{n+1}.
- Note that on Δ_{n+1} the equality $R(x_0, \ldots, x_n).(x_0 + x_1 + \ldots + x_n)^N = R(x_0, \ldots, x_n)$ holds.
- As a note of caution consider for example the polynomial $(x_0 - x_1)^2 \geq 0$. However $(x_0 - x_1)^2(x_0 + x_1)^N$ will have at least one negative coefficient for any positive integer N because the value of the expression is zero at the point $(\frac{1}{2}, \frac{1}{2})$.
- The coefficients of $R(x_0, \ldots, x_n).(x_0 + x_1 + \ldots + x_n)^N$ can be recursively computed for $N = 1, 2, 3, \ldots$, using a "Pascal-triangle"-type procedure. (This is explained in [4]).
- Note that due to the homogeneity of R this polynomial will be nonnegative on the simplex if and only if it is nonnegative on the orthant $\mathbb{R}^{n+1}_{+}$. To investigate whether it is non-negative on any other orthant the transformation $x_i \mapsto -x_i, \ i \in I$ for the appropriate index set I associated to the particular orthant, can be applied, to map that orthant to the positive orthant. Then our methods can be applied to that transformed problem. This means that in total 2^n cases will have to be considered in general.

- A further note of caution relates to the fact that if a non-homogeneous polynomial p which is (strictly) positive on the positive orthant is homogenized, the corresponding homogenized polynomial $P(x_0, x_1, \ldots, x_n) = x_0^{\deg(p)}.p(x_1/x_0, x_2/x_0, \ldots, x_n/x_0)$ does *not* have to be strictly positive on the simplex Δ_{n+1}. An example of this is given by $p(x_1, x_2) = x_1^2 + x_2$, which has corresponding homogenized polynomial $P(x_0, x_1, x_2) = x_1^2 + x_2.x_0$. Clearly $p > 0$ on the positive orthant $(0, \infty) \times (0, \infty)$, while P is zero in the point $(0, 0, 1)$ on the simplex Δ_{n+1}.
- Using results of [5] the Polya test can be applied to determine whether for any candidate value of γ the minimum of $P - \gamma.Q$ is negative on the simplex Δ_{n+1}. In this way the infimum can be determined in a finite, but possibly large, number of steps. Further details about this will be explained in the presentation.

7. Concluding remarks. Some examples will be presented to show how the methodology presented works. Until now we only have been able to run rather small examples, due to memory and computing time requirements involved in these computations. Some remarks will be made about possible ways to get faster and more efficient computer implementations. One interesting question is whether the combined usage of the "upper bound" methods (MC, QMC and other relevant optimization algorithms) with the "lower bound" methods (Polya method; one could also think of so-called sum-of-squares methods) could lead to algorithms that are faster than those which just concentrate on one of these possibilities. It may be good to add that for the moment it is envisaged that the methods developed here could be of help to test existing optimization algorithms, to see if they produce a global optimum in, perhaps relatively small, test cases. Applications to model reduction problems is one area where this could be of interest.

REFERENCES

[1] B. Hanzon, D. Jibetean "Global Minimization of Multivariate Polynomial Using Matrix Methods". vol 27, pp. 1-23, 2001. *J. Global Optimization,* Vol 27, pp. 1-23, 2001.

[2] D. Jibetean "Global Optimization of Rational Multivariate Functions" *CWI Report PNA-R0120,* October 31, 2001, CWI, Amsterdam Online version available at http://oai.cwi.nl/oai/asset/4299/04299D.pdf

[3] H. Niederreiter "Random Number Generation and Quasi-Monte Carlo Methods" *SIAM, PA, USA,* 1992

[4] B. Hanzon "Mathematics from the Past, Mathematics for the Future"; *Inaugural Lecture,* 12 Dec 2012, University College Cork An online link to the lecture can be found on euclid.ucc.ie/pages/staff/hanzon/hanzon.m

[5] V. Powers, B. Reznick "A new bound for Polya's Theorem with applications to polynomials positive on polyhedra" *J. Pure Appl. Algebra* Vol. 164, 2001