

A Model Problem for Global optimization

Siegfried M. Rump^{1a)}

¹ *Institute for Reliable Computing, Hamburg University of Technology, Schwarzenbergstraße 95, Hamburg 21071, Germany, and Visiting Professor at Waseda University, Faculty of Science and Engineering, 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan, and Visiting Professor at Université Pierre et Marie Curie (Paris 6), Laboratoire LIP6, Département Calcul Scientifique, 4 place Jussieu, 75252 Paris cedex 05, France.*

^{a)} rump@tu-harburg.de

[Received January 11, 2010/Revised January 11, 2010/Published January 11, 2010]

[DOI: 10.1588/nolta.1.1]

Abstract : We present a model problem for global optimization in a specified number of unknowns. We give constraint and unconstraint formulations. The problem arose from structured condition numbers for linear systems of equations with Toeplitz matrix. We present a simple algorithm using additional information on the problem to find local minimizers which presumably are global. Without this it seems quite hard to find the global minimum numerically. For dimensions up to $n = 18$ rigorous lower bounds for the problem are presented.

Key Words : global optimization, model problem, structured perturbations, Toeplitz matrix, INTLAB

1. Problem formulation

Let $P, Q \in \mathbb{R}[x]$ be real polynomials. For $P(x) = \sum_{\nu=0}^{n-1} p_{\nu}x^{\nu}$ define

$$\|P\| := \|p\|_2 = \left(\sum_{\nu=0}^{n-1} p_{\nu}^2 \right)^{1/2}. \quad (1)$$

For given $2 \leq n \in \mathbb{N}$ solve

$$\|PQ\| = \min! \quad \text{subject to} \quad \deg(P), \deg(Q) \leq n-1 \text{ and } \|P\| = \|Q\| = 1, \quad (2)$$

where PQ denotes the polynomial multiplication (convolution). This is equivalent to the computation of

$$\mu_n := \min\{\|PQ\|^2 : \deg(P), \deg(Q) \leq n-1 \text{ and } \|P\|^2 = \|Q\|^2 = 1\}, \quad (3)$$

a constraint optimization problem in $2n$ unknowns. Scaling the factors by a power of x does not change the norms, so the unconstrained problem

$$\mu_n := \min \left\{ \frac{\|PQ\|^2}{\|P\|^2\|Q\|^2} : \deg(P) = \deg(Q) = n - 1 \text{ and } p_{n-1} = q_{n-1} = 1 \right\} \quad (4)$$

in $2n - 2$ unknowns can be solved as well. Note that the problem arose in matrix theory (see below), so μ_n refers to polynomials of degree $n - 1$.

We mention that for other norms this problem is solved. For example, when replacing the norm in (3) or (4) by the maximum modulus on the interval $[-1, 1]$, Kneser [10] computed the minimum explicitly and showed that the Chebyshev polynomials are the minimizers. If the norm is the maximum modulus of the polynomial on the unit circle the global minimum is also explicitly known [2], [3]. Moreover, the problem is solved in [1] for the Bombieri 2-norm $[P]_2$ defined by

$$[P]_2 := \left[\sum_{\nu=0}^{n-1} \binom{n-1}{\nu} p_\nu^2 \right]^{1/2}.$$

We have a very efficient algorithm to compute an upper bound for μ_n , and we strongly believe that our approximations are very close to the true minimum. The challenge is to compute rigorous *lower bounds* for μ_n . The best known lower and upper bounds for μ_n are displayed in Table I. The lower bounds are correctly rounded [9], and we believe that the upper bounds are correct to the last digit.

Table I. Known lower and upper bounds for μ_n .

n	lower bound	upper bound	β_1	β_2
2	0.5	0.5	5.00e-1	1.29e-1
3	1/9	1/9	4.17e-2	7.54e-3
4	1.742917332e-2	1.74291733214326528802e-2	5.56e-3	5.22e-4
5	2.339595548e-3	2.33959554815559112417e-3	3.97e-4	3.94e-5
6	2.897318752e-4	2.89731875279681924918e-4	3.97e-5	3.13e-6
7	3.41850698e-5	3.41850698000828404075e-5	2.71e-6	2.57e-7
8	3.9054356e-6	3.90543564975572493986e-6	2.38e-7	2.16e-8
9	4.360016539181e-7	4.36001653918104846130e-7	1.59e-8	1.85e-9
10	4.783939568e-8	4.78393956877097593268e-8	1.30e-9	1.61e-10
11	5.1787e-9	5.17874909744699053306e-9	8.52e-11	1.41e-11
12	5.54539e-10	5.54588183116313476112e-10	6.64e-12	1.25e-12
13	5.881019273e-11	5.88668808118660931300e-11	4.33e-13	1.12e-13
14	6.1e-12	6.20244499205390502197e-12	3.27e-14	1.00e-14
15	6.0e-13	6.49436541858095128797e-13	2.12e-15	9.08e-16
16	6.0e-14	6.76360425582213790573e-14	1.56e-16	8.24e-17
17	6.0e-15	7.01126319701437415845e-15	1.00e-17	7.51e-18
18	1.0e-16	7.23839447969438758614e-16	7.25e-19	6.87e-19
19	?	7.44600395577761355965e-17	4.66e-20	6.31e-20
20	?	7.63505369377296099185e-18	3.32e-21	5.81e-21
21	?	7.80646424601126297071e-19	2.13e-22	5.36e-22
22	?	7.96111666797470505639e-20	1.49e-23	4.96e-23
23	?	8.09985434822450864918e-21	9.55e-25	4.60e-24
24	?	8.22348468914618752033e-22	6.64e-26	4.28e-25
25	?	8.33278066679120912785e-23	4.24e-27	3.98e-26

2. Mathematical background

Let a linear system $Ax = b$ be given. The normwise condition number is defined by

$$\kappa(A, x) := \limsup_{\epsilon \rightarrow 0} \left\{ \frac{\|\Delta x\|}{\epsilon \|x\|} : (A + \Delta A)(x + \Delta x) = b + \Delta b, \Delta A \in M_n(\mathbb{R}), \Delta b \in \mathbb{R}^n, \|\Delta A\| \leq \epsilon \|A\|, \|\Delta b\| \leq \epsilon \|b\| \right\}. \quad (5)$$

It is well-known [7] that

$$\kappa(A, x) = \|A^{-1}\| \|A\| + \frac{\|A^{-1}\| \|b\|}{\|x\|}, \quad (6)$$

so that $b = Ax$ implies

$$\|A^{-1}\| \|A\| \leq \kappa(A, x) \leq 2\|A^{-1}\| \|A\|.$$

Let a linear system $Ax = b$ with Toeplitz matrix A be given. A specialized solver needs only the first row and column of A as input. Thus general perturbations are not possible, only Toeplitz perturbations. Accordingly, the sensitivity of the system should be judged by the Toeplitz condition number, not by the general condition number. The Toeplitz condition number $\kappa^{\text{Toep}}(A, x)$ is naturally defined by restricting the perturbations ΔA in (5) to Toeplitz matrices. The question arises how small the ratio between the structured and unstructured condition number can be.

In [14] we proved that this ratio satisfies

$$\frac{\kappa^{\text{Toep}}(A, x)}{\kappa(A, x)} = \alpha \frac{\|A^{-1}J\Psi_x\|}{\|A^{-1}\| \|x\|} \geq \frac{1}{\sqrt{n}} \frac{\sigma_{\min}(\Psi_x)}{\|x\|},$$

where $\frac{1}{\sqrt{n}} \leq \alpha \leq \sqrt{2}$ and the matrix Ψ_x is defined by

$$\Psi_x := \begin{pmatrix} x_1 & x_2 & \dots & x_n & & \\ & x_1 & x_2 & \dots & x_n & \\ & & & \dots & & \\ & & & & x_1 & x_2 & \dots & x_n \end{pmatrix} \in \mathbb{R}^{n \times (2n-1)}, \quad (7)$$

where $J \in \mathbb{R}^{n \times n}$ is the permutation matrix mapping $(1, \dots, n)^T$ into $(n, \dots, 1)^T$.

Surprisingly, the lower bound depends only on the solution x , not on the matrix A . However, no closed formula for the minimum ratio was known, and it was also not clear for some time whether actually Toeplitz matrices exist realizing anything close to the lower bound. This was solved in [15]. We proved that the minimum ratio of the condition numbers satisfies

$$\sqrt{2\mu_n} \geq \inf \left\{ \frac{\kappa^{\text{Toep}}(A, x)}{\kappa(A, x)} : A \in \mathbb{R}^{n \times n} \text{ Toeplitz}, 0 \neq x \in \mathbb{R}^n \right\} \geq \sqrt{\frac{\mu_n}{n}}$$

for all n . Moreover, the μ_n defined in (3) or (4) are by

$$\sqrt{\mu_n} = \min_{\|x\|=1} \sigma_{\min}(\Psi_x) > 0 \quad (8)$$

directly related to the smallest possible singular value of Ψ_x defined in (7).

From this interpretation a simple algorithm [15] can be derived to approximate μ_n . Denote for any $0 \neq x \in \mathbb{R}^n$ the left singular vector of Ψ_x to the smallest singular value $\sigma_{\min}(\Psi_x)$ by y . Then the entries of

$$y^T \Psi_x$$

are the coefficients of the polynomial $x(t)y(t)$, where $x(t)$ and $y(t)$ are the polynomials with coefficients x and y , respectively. By construction we have $\|y^T \Psi_x\| = \sigma_{\min}(\Psi_x)$. Since polynomial multiplication is commutative it follows

$$y^T \Psi_x = x^T \Psi_y \quad \text{and} \quad \sigma_{\min}(\Psi_y) \leq \|x^T \Psi_y\|.$$

Replacing x by y we calculate the left singular vector y of the new Ψ_x and so forth generating a monotonically decreasing sequence of upper bounds for μ_n^2 . The following Matlab [11] Algorithm 1 implements this method.

Algorithm 1 Approximation of μ_n .

```

function [mu_n,P,Q] = globoptmin(n)
% Approximate local minimum and minimizers for mu_n. Call
%   [mu_n,P,Q] = globoptmin(n)
%
z = zeros(1,n-1);
p = poly(randn(1,n-1));
r = p/norm(p);
minsvd = 1;
minsvdold = 0;
while abs((minsvd-minsvdold)/minsvd)>1e-14
    A = toeplitz([r(1) z],[r z]);
    [U S V] = svd(A);
    r = U(:,n)';
    minsvdold = minsvd;
    minsvd = S(n,n);
end
mu_n = minsvd^2;
P = A(1,1:n);
Q = r;

```

The algorithm uses a random starting vector and converges almost always to the same value. This value calculated with some multiple precision package is displayed (correctly rounded) as upper bound in Table I. The computation takes only few seconds. Supposedly this is the global minimum; for smaller dimensions this heuristic is verified by Table I.

3. Simplification of the problem and known lower bounds

A significant simplification of the problem is that we may assume without loss of generality in (3) that both P and Q have all roots on the unit circle. More precisely it is shown in [15] that for a fixed polynomial P of degree $n - 1$ the minimum of

$$\min\{\|PQ\| : \deg(Q) \leq n - 1 \text{ and } \|Q\| = 1\} \quad (9)$$

is achieved by a polynomial Q having all roots on the unit circle. It was noted in [5] that the so called Caratheodory-representation of the autocorrelation Toeplitz matrix $\Psi_p \Psi_p^T$ (deduced from a theorem of Caratheodory [6], Theorem 4.1) even implies that Q can be found such that all roots of Q are simple.

It follows that for polynomials P, Q realizing μ_n that the coefficient vectors may be assumed to be symmetric or skewsymmetric, i.e. $p_{n-1-\nu} = \pm p_\nu$ and similarly for Q . Thus (3) can be rewritten into three optimization problems with about half the number of unknowns, where the smallest of the three minima is equal to μ_n . We call this the “modified problem”.

Another possibility to attack the problem is to verify a guessed lower bound α to be a true lower bound by showing

$$\min\{\|PQ\|^2 - \alpha\|P\|^2\|Q\|^2 : \deg(P) = \deg(Q) = n - 1 \text{ and } p_{n-1} = q_{n-1} = 1\} \geq 0. \quad (10)$$

If (10) is true it obviously implies $\mu_n \geq \alpha$.

The lower bounds in Table I are taken from a paper by Kaltofen et al. [9] improving the results in [8]. He transforms the problem into a semidefinite programming problem and shows that a certain function is a sum of squares. For higher values of n the problem becomes increasingly ill-conditioned and multiple precision arithmetic is used. To prove $\mu_{16} \geq 6 \cdot 10^{-14}$ they needed a little more than 1 day of computing time, to prove $\mu_{17} \geq 6 \cdot 10^{-15}$ more than 18 days. The much weaker bound $\mu_{18} \geq 1 \cdot 10^{-16}$ needed 25 days of computing time.

Other optimization packages usually find a verified lower bound up to dimension $n \leq 4$ in reasonable computing time. An exception is Martin Berz's COSY package (thanks to Kyoko Makino), which could compute bounds up to $n \leq 8$ for the modified problem.

Using (10) the problem can be rewritten into a sums-of-squares problem by SOSTOOLS [12]. The problem can be reduced to verify positive semi-definiteness of some matrix W . An inclusion of this matrix $[W]$ can be computed in double precision interval arithmetic and by INTLAB [13], the Matlab toolbox for Reliable Computing. Using algorithm "isspd" in INTLAB one can verify positive semi-definiteness of all matrices within this interval matrix $[W]$, in particular of W . This approach verifies about 2 correct digits of the lower bound up to $n \leq 12$. Note that only double precision and no multiple precision arithmetic is used.

Another approach is to attack the problem by using Gröbner bases. Mohab Safey El Din [16] from Paris VI reduced the problem to finding the smallest positive root of a univariate polynomial $G_{[n]}$, the degrees and coefficients of which are as in Table II.

Table II. Polynomials related to Gröbner bases.

n	degree($G_{[n]}$)	coeff($G_{[n]}$)
5	18	~ 20 digits
6	35	~ 50 digits
7	131	~ 170 digits
8	255	~ 430 digits

Moreover there are some general lower bounds. Using Proposition 1.B.4 with $m = n$ in [1] one can show [4]

$$\mu_n \geq \beta_1 := \binom{2n-2}{n-1}^{-1} \binom{n-1}{\lfloor (n-1)/2 \rfloor}^{-2}, \quad (11)$$

and Theorem 2.9 in [15] gives

$$\mu_n \geq \beta_2 := \frac{4}{(2n-1) \cdot \Delta^{2n-2}}, \quad (12)$$

where $\Delta := e^{4G/\pi} = 3.209912\dots$ for Catalan's constant $G = 0.915965\dots$. The results are displayed in Table I. As can be seen the first bound is best for small values of n , whereas for larger values and asymptotically the second one is the best. There is also a factor coefficient bound $\mu_n \geq \binom{2n-2}{n-1}^{-2}$ by Mignotte, which is, of course, always worse than β_1 .

For convenience, on our homepage <http://www.ti3.tu-harburg.de/rump> we give AMPL-like problem formulations for the original problem and for the modified problem using the symmetry of the coefficients. Moreover we give local minimizers which we believe are close to the global ones. Those are computed by Algorithm 1.

Vast computational experience led us in [15] to the following conjecture.

Conjecture. For all n there exist unique minimizing monic polynomials P, Q of degree $n-1$ for (3) such that all coefficients of P are positive, and $Q(x) = (-1)^{n-1}P(-x)$. The roots $a_\nu \pm ib_\nu$ of P have all positive real parts, and the roots of Q are $-a_\nu \pm ib_\nu$.

The computed local minimizers (in double precision) display this property. A proof for that would reduce the number of unknowns again significantly.

Quite a number of very interesting results on this optimization problem can be found in [5]. Bünger gives arguments supporting that there is a unique minimizer for the real problem, that the complex and real optimization problem have the same minimum and, up to scaling, the same minimizer. Moreover he gives much simpler proofs for some results in [15]. His work endorses our conjecture above.

References

- [1] B. Beauzamy, E. Bombieri, P. Enflo, and H. Montgomery. Products of Polynomials in Many Variables. *Journal of Number Theory*, 36:219–245, 1990.

- [2] D.W. Boyd. Two sharp inequalities for the norm of a factor of a polynomial. *Mathematika*, 39:341–349, 1992.
- [3] D.W. Boyd. Sharp Inequalities for the Product of Polynomials. *Bull. London Math. Soc.*, 26:449–454, 1994.
- [4] F. Bünger. private communication, 2008.
- [5] F. Bünger. The euclidean norm of the product of two polynomials. manuscript, 2008.
- [6] U. Grenander and G. Szegö. *Toeplitz forms and Their applications*. Univ. California Press, Berkeley, 1958.
- [7] N.J. Higham. *Accuracy and Stability of Numerical Algorithms*. SIAM Publications, Philadelphia, 2nd edition, 2002.
- [8] E. Kaltofen, B. Li, Z. Yang, and L. Zhi. Exact certification of global optimality of approximate factorizations via rationalizing sums-of-squares with floating point scalars. *ISSAC*, pages 155–163, 2008.
- [9] E. Kaltofen, B. Li, Z. Yang, and L. Zhi. Exact certification in global polynomial optimization via sums-of-squares of rational functions with rational coefficients. submitted for publication, 2009.
- [10] H. Kneser. Das Maximum des Produkts zweier Polynome. In *Sitz. Preuss. Akad. Wiss., Phys.-Math. Kl.*, pages 426–431, 1934.
- [11] MATLAB User’s Guide, Version 7. The MathWorks Inc., 2004.
- [12] S. Prajna, A. Papachristodoulou, P. Seiler, and P. A. Parrilo. *SOSTOOLS: Sum of squares optimization toolbox for MATLAB*, 2004.
- [13] S.M. Rump. INTLAB - INTerval LABoratory. In Tibor Csendes, editor, *Developments in Reliable Computing*, pages 77–104. Kluwer Academic Publishers, Dordrecht, 1999.
- [14] S.M. Rump. Structured Perturbations Part I: Normwise Distances. *SIAM J. Matrix Anal. Appl. (SIMAX)*, 25(1):1–30, 2003.
- [15] S.M. Rump and H. Sekigawa. The ratio between the Toeplitz and the unstructured condition number. *Operator Theory: Advances and Applications*, 199:397–419, 2009.
- [16] M. Safey. private communication, 2009.