Control and Cybernetics

vol. **39** (2010) No. 4

The EM Algorithm applied to determining new limit points of Mahler measures*

by

Souad El Otmani, Georges Rhin and Jean-Marc Sac-Épée

Université Paul Verlaine-Metz, LMAM, CNRS UMR 7122,
Ile du Saulcy, METZ, F-57045, France
e-mail: elotmani@univ-metz.fr, rhin@univ-metz.fr,jmse@univ-metz.fr

Abstract: In this work, we propose new candidates expected to be limit points of Mahler measures of polynomials. The tool we use for determining these candidates is the Expectation-Maximization algorithm, whose goal is to optimize the likelihood for the given data points, i.e. the known Mahler measures up to degree 44, to be generated by a specific mixture of Gaussians. We will give the mean (which is a candidate to be a new limit point) and the relative amplitude of each component of the more likely gaussian mixture.

Keywords: Mahler Measure, EM algorithm.

1. State of the art

Recall that if P is a polynomial defined as

$$P(x) = \sum_{k=0}^{n} a_k x^k, a_k \in \mathbb{C} \text{ and } a_n \neq 0,$$

then its Mahler measure (see Boyd, 1980, 1989; Mossinghoff, 1998) is defined to be

$$M(P) = |a_n| \prod_{k=1}^{n} max(1, |\alpha_k|),$$

where the α_k's are the roots of P.

Initiated by Lehmer (1933) who, for polynomials P with *integer coefficients*, provided the smallest values of $M(P)$ for $\deg(P) = 2,\ 3,$ and 4, and for reciprocal P with $\deg(P) = 2,\ 4,\ 6$ and 8, computations on Mahler measure were continued by Boyd (1980, 1981, 1989) who found all reciprocal P with $M(P) \leq 1.3$

*Submitted: July 2010; Accepted: October 2010.

and degree up to 20, as well as those with $M(P) \leq 1.3$ and degree up to 32 with height 1. Boyd's lists were extended by Mossinghoff (1998) up to degree 24 (up to degree 40 for height 1 polynomials). Flammang, Grandcolas and Rhin (1999) proved the completeness of these lists, and Flammang, Rhin and Sac-Épée (2006) provided all polynomials P such that $M(P) \leq \theta_0$ ($\theta_0 \simeq 1.3247\ldots$ is the smallest Pisot number) with $\deg(P) \leq 36$, and all polynomials P such that $M(P) \leq 1.31$ and $\deg(P) \leq 40$.

On Mossinghoff's web site (see *http://www.cecm.sfu.ca/ mjm/Lehmer/lists*), we can find a list of all known noncyclotomic and irreducible polynomials with integer coefficients and degree at most 180 and Mahler measure below 1.3, including polynomials provided by P. Lisonek (2000), G. Rhin and J.-M. Sac-Epée (2003), and Mossinghoff, Rhin, and Wu (2008) who proved the completeness of the list up to degree 44. For each polynomial, its Mahler measure is available.

One of the important questions concerning the Mahler measure is the following: What are the *small* limit points of the Mahler measure of the set of algebraic integers ? In a recent paper, Boyd and Mossinghoff (2005) gave a list of 48 such limit points less than 1.37. We note that there are only two limit points less than 1.3, namely $1.255\ldots$ and $1.285\ldots$ All are obtained by values of Mahler measures of polynomials in several variables of different types. We may add the following question known as Lehmer's problem: Is 1 a limit point of the set of the Mahler measures? The smallest value of the Mahler measure that is known is $1.176\ldots$ given by the polynomial $X^{10}+X^9-X^7-X^6-X^5-X^4-X^3+X+1$, found by D.H. Lehmer himself (Lehmer, 1933).

For a more complete survey, please refer to Smyth (2008).

Then, an interesting idea is to suggest a new statistical approach, which is conceivable having regard to complete lists of values which are available up to degree 44, and very rich lists up to degree 180. So, we will focus on the possible values of limit points smaller than 1.3 using these tables as incoming data, and the EM algorithm as a statistical analysis tool to suggest the existence of two possible new values of limit points as 1.256533... and 1.286625...

In the following section, we draw the histogram of the 8415 available points with the purpose of investigating what kind of distributions this list of points could arise from.

2. Graphical observations

Let us examine very closely the histogram of the frequency distribution (Fig. 1) of the given points, with zooms at interesting zones.

By zooming around the first known limit point $1.255433866\ldots$, we obtain Fig. 2.

The peculiar appearance of the graphic induces us to speculate that the distribution of scalar values around the first known limit point does not follow a simple gaussian model, but rather arises from a mixture gaussian model, as in the following example (see Fig. 3) corresponding to a bimodal mixture made of

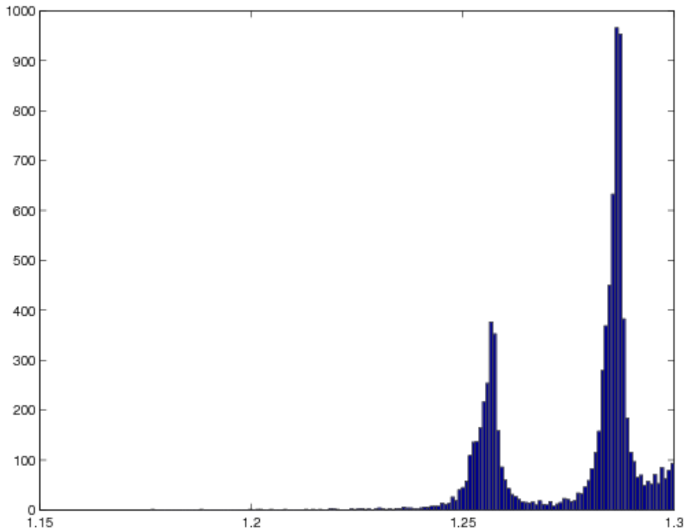

Figure 1. Histogram of the frequency distribution of the known limit points

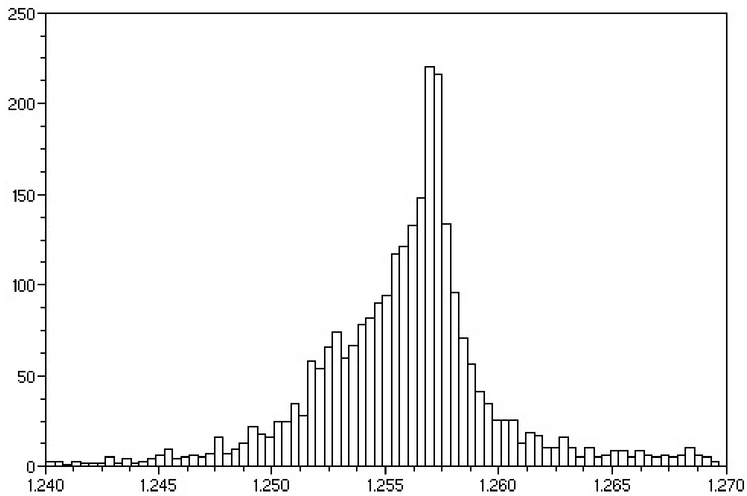

Figure 2. Zoom around the first known limit point

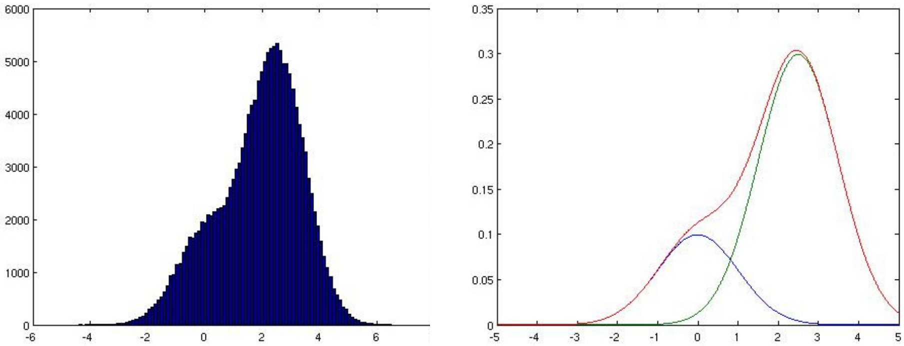

Figure 3. Example corresponding to a bimodal gaussian

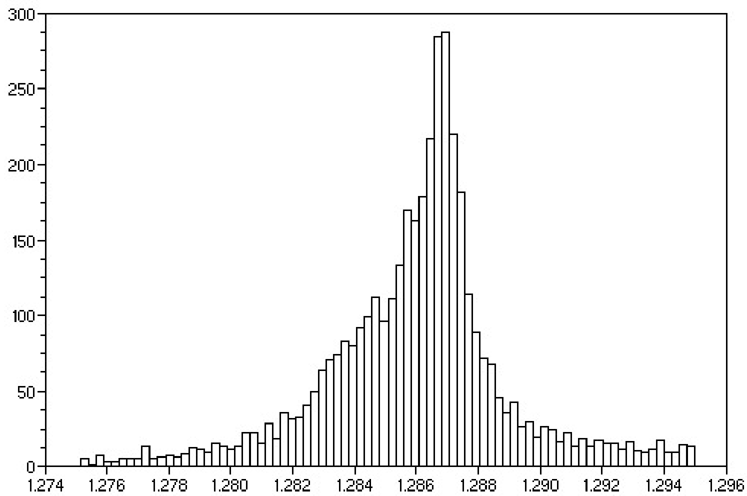

Figure 4. Zoom around the second known limit point

4000 points issued from the $\mathcal{N}(0,1)$ Gaussian law and 12000 points issued from the $\mathcal{N}(2.5,1)$ Gaussian law.

The same observation can be made when zooming around the second known limit point $1.285734864\ldots$ (see Fig. 4).

For determining the parameters of the underlying gaussian mixture, we use a systematic method consisting in applying the EM algorithm (see Dasgupta and Schulman, 2000; MacLachlan and Krishnan, 2008; Tagare, 1998; Xu and Jordan, 1996; Wu, 1983) to the given list of points.

For the reader's convenience, we outline some EM theory in Section 3.

3. The EM algorithm

Consider n independent scalar values $a_1, a_2, \cdots, a_n$. Each a_i is supposed to arise from a probability distribution whose density can be expressed as

$$f(x \mid \theta) = \sum_{j=1}^{N} p_j g_j(x \mid \mu_j, \sigma_j).$$

Scalar value p_j stands for the mixing proportion of the jth component of the mixture, and we have $\sum_{j=1}^{N} p_j = 1$ and $\forall j = 1, \cdots, N, \quad 0 < p_j < 1$.
Function $g_j(\cdot \mid \mu_k, \sigma_k)$ is the gaussian density with mean μ_j and standard deviation σ_j, and is defined by

$$g_j(x \mid \mu_j, \sigma_j) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_j)^2}{2\sigma_j^2}}.$$

$\theta = (p_1, \cdots, p_{N-1}, \mu_1, \cdots, \mu_N, \sigma_1, \cdots, \sigma_N)$ is a vector whose components are the mixture parameters, which are estimated by maximizing the loglikelihood

$$L(\theta \mid a_1, \cdots, a_n) = \sum_{i=1}^{n} \ln\Big(\sum_{j=1}^{N} p_j g(a_i \mid \mu_j, \sigma_j)\Big).$$

Given vector θ, the belonging $h(k, l)$ of data point a_k to cluster number l can be computed by using Bayes' theorem as

$$h(k, l) = p(cluster's\ number = l \mid a_k, \theta) = \frac{p_l g(x_k \mid \mu_l, \sigma_l)}{\sum_{i=1}^{N} p_i g(x_k \mid \mu_i, \sigma_i)}.$$

One of the most liked method used for determining the maximum likelihood solution is the Expectation-Maximization algorithm. Roughly speaking, assuming that given data arise from a gaussian mixture model with N components, the EM algorithm is devoted to estimate the parameters (means, standard deviations) of each component of the mixture for which the observed data are the most likely.

The Expectation-Maximization algorithm for gaussian mixtures is an iterative process defined as follows:

- Choose initial parameters settings.

- Repeat until convergence:
  - **E-Step**: Using the current parameter values, compute $h(k,l)$ for $1 \le k \le n, \quad 1 \le l \le N$:

$$h(k,l)^{(i)} = \frac{p_l^{(i)} g(x_k \mid \mu_l^{(i)}, \sigma_l^{(i)})}{\sum_{i=1}^{N} p_l^{(i)} g(x_k \mid \mu_l^{(i)}, \sigma_l^{(i)})},$$

  - **M-Step**: Use data points a_k and just computed values $h(k,l)$ to give new parameters values:
    * $S_l^{(i+1)} = \sum_{k=1}^{N} h(k,l)^{(i)}$
    * $\alpha_l^{(i+1)} = \frac{1}{N} S_l^{(i+1)}$
    * $\mu_l^{(i+1)} = \frac{1}{S_l^{(i+1)}} \sum_{k=1}^{N} h(k,l)^{(i)} a_k$
    * $(\sigma_l^{(i+1)})^2 = \frac{1}{S_l^{(i+1)}} \sum_{k=1}^{N} h(k,l)^{(i)} (a_k - \mu_l^{(i+1)})^2.$

We stop the iterative process when the log-likelyhood's value becomes almost unchanged from one iteration to the next.

4. Application for search of candidates to be new limit points

Many softwares implementing EM algorithm are available on the web. Among all, we choose to use Mixmod Software, which is an exploratory data analysis tool for solving clustering and classification problems.

A careful observation of histograms above induced us to surmise that the given list of points arises from a gaussian mixture constituted by four components that we plan to make more precise.

Our choice was to work with Mixmod in Scilab environment.

Around the first known limit point $1.255433866\ldots$, we applied EM algorithm on interval $(1.24, 1.27)$. Results obtained after calculations are summarized in the following table:

Table 1. Two clusters on the first interval

	means	proportions
cluster 1	1.256533	0.433147
cluster 2	1.255336	0.566853

Around the second known limit point 1.285734864 . . . , we applied EM algorithm on interval $(1.275, 1.295)$. Results obtained after calculations are summarized in the following table:

Table 2. Two clusters on the second interval

	means	proportions
cluster 1	1.286625	0.36126
cluster 2	1.285674	0.63874

On each interval, calculations provided a precise approximation of the already known limit value, and a new value expected to be a new limit point for Mahler measures of polynomials. So, our two new candidates are 1.256533 and 1.286625.

5. Conclusion

While these two new values seem to be promising, one should keep in mind that contrary to known limit values, these new values are not mathematically proved to be limit points. Numerical investigations simply lead us to consider these points as good candidates, worthy of some more detailed theoretical studies.

References

Boyd, D.W. (1981) Speculations concerning the range of Mahler's measure. *Canad. Math. Bull.* **24**, 453-469.

Boyd, D.W. (1980) Reciprocal polynomials having small measure. *Math. Comp.* **35**, 1361-1377.

Boyd, D.W. (1989) Reciprocal polynomials having small measure II. *Math. Comp.* **53**, 355-357, S1-S5.

Boyd, D.W. (2005) M. J. Mossinghoff, Small Limit Points of Mahler's Measure. *Experimental Mathematics* **14**, Part 4, 403-414.

Dasgupta, S. and Schulman, L.J. (2000) A Two-Round Variant of EM for Gaussian Mixtures. *16th Conference on Uncertainty in Artificial Intelligence*, 152-159.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum Likelihood from Incomplete Data via the EM Algorithm. *J. Royal Statist. Soc. Ser. B.*, **39**, 1-38.

Flammang, V., Grandcolas, M. and Rhin, G. (1999) Small Salem numbers. *Number Theory in Progress*, 1 (Zakopane-Kościelisko, 1997), de Gruyter, Berlin, 165-168.

Flammang, V., Rhin, G. and Sac-Épée, J.M. (2006) Integer Transfinite Diameter and polynomials with small Mahler measure. *Math. Comp.* **75**, 1527-1540.

LEHMER, D.H. (1933) Factorization of certain cyclotomic functions. *Ann. of Math.* **2** (34), 461-479.

MCLACHLAN, G.J. and KRISHNAN, T. (2008) *The EM Algorithm and Extensions. Wiley Series in Probability and Statistics (Second, ed.).* Wiley-Interscience, Hoboken, NJ.

MOSSINGHOFF, M.J. (1998) Polynomials with small Mahler measure. *Math. Comp.* **67**, 1697-1705.

MOSSINGHOFF, M.J., RHIN, G. and WU, Q. (2008) Minimal Mahler Measures. *Experiment. Math.*, **17** (4), 451-458.

RHIN, G. and SAC-ÉPÉE, J.M. (2003) New methods providing high degree polynomials with small Mahler measure. *Experiment. Math.*, **12** (4), 457-461.

TAGARE, H.D. (1998) *A Gentle Introduction to the EM Algorithm, Part I: Theory*. Available on the web.

XU, L. and JORDAN, M.I. (1996) On Convergence Properties of the EM Algorithm for Gaussian Mixtures. *Neural Computation* **8**, 129-151.

WU, C.F.J. (1983) On the Convergence Properties of the EM Algorithm. *The Annals of Statistics* **11** (1), 95-103.

SMYTH, C. (2008) The Mahler measure of algebraic numbers: A survey. *Number Theory and Polynomials*, London Math. Soc. Lecture Note Ser. 352, Cambridge Univ. Press, 322-349.