THE NUMBER OF TREES HALF OF WHOSE VERTICES ARE LEAVES AND ASYMPTOTIC ENUMERATION OF PLANE REAL ALGEBRAIC CURVES

V.M. Kharlamov, S.Yu. Orevkov

Abstract. The number of topologically different plane real algebraic curves of a given degree d has the form $\exp(Cd^2 + o(d^2))$. We determine the best available upper bound for the constant C. This bound follows from Arnold inequalities on the number of empty ovals. To evaluate its rate we show its equivalence with the rate of growth of the number of trees half of whose vertices are leaves and evaluate the latter rate.

Introduction

0.1. Plane projective curves and rooted trees. Recall that a *rooted tree* is a tree with a distinguished vertex. The distinguished vertex is called the *root.* The *multiplicity* or the *valence* of a vertex is the number of edges which are incident to it. A vertex of multiplicity one is called a *leaf.* By convention, we assume that the root is a leaf if the tree has no other vertices. Otherwise, the root is not considered as a leaf even if its multiplicity is one. The vertices of multiplicity > 1 are called *internal.*

In this paper we work exclusively with *unlabeled finite trees* and use them to encode the topology of nonsingular curves in the real projective plane.

By a nonsingular curve we mean a closed one-dimensional, not necessarily connected, compact sub-manifold. Each connected component of such a curve is a topological circle smoothly embedded in $\mathbb{R}P^2$. There are two species of embedded circles: *one-sided* circles, which, similar to a projective line, do not decompose $\mathbb{R}P^2$, and *two-sided* circles, which, similar to a standard circle, decompose $\mathbb{R}P^2$ in a disc and a Moebius band. Following the real algebraic geometry tradition, the two-sided components are called *ovals* even though they may be nonconvex. The number of one-sided components is at most one. By analogy with the algebraic case (see Section 0.2), if all the curve components are ovals, we say that the curve is of *even degree*, and otherwise, that it is of *odd degree.*

To encode the topology of a curve we prefer to use the connected components of the complement of the curve. If the degree is even, one of the components of the complement is nonorientable and the other components are orientable as well as their closures. If the degree is odd, all the components of the complement are orientable but the closure of one and only one of them is nonorientable (this is the complement component adjacent to the one-sided component of the curve).

First author is a member of Research Training Networks EDGE and RAAG, supported by the European Human Potential Program.

Finally, our encoding will look as follows. We associate the vertices with the connected components of the complement of the curve. The root will correspond to the component with non-oriented closure and the tree will represent the adjacency relations between the components (see Figure 1). The fact that this graph is a tree follows from the Jordan curve theorem. It is **finite** since our curves are compact. The number of edges is equal to the number of ovals, so that the number of vertices is the same as the number of components of the curve if the degree is odd, and it is greater by 1 if the degree is even.

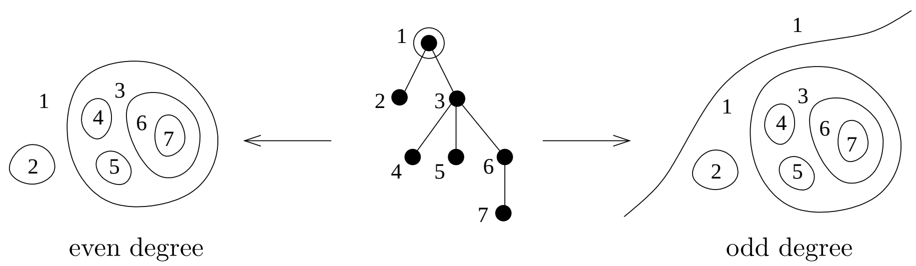

FIGURE 1. ROOTED TREE AND THE CORRESPONDING PLANE CURVES

Two curves have the same encoding and the same degree parity if and only if there is an ambient isotopy transforming one into another, so that these two invariants, the tree and the degree parity, describe completely the isotopy class of a curve. In classical terminology, the isotopy classes are called *arrangements*. (If one likes, he can speak of ambient homeomorphisms and ambient homeomorphism classes instead of isotopies and isotopy classes; in the case of curves in $\mathbb{R}P^2$ it is an equivalent setting.)

Let us notice that the ovals corresponding to leaves are called *empty ovals*.

0.2. Statement of results. In this paper we are interested in algebraic nonsingular curves. More precisely, a nonsingular algebraic (real plane) curve of degree d is a curve given, in homogeneous coordinates, by a polynomial equation $p(x, y, z) = 0$, where p is a real homogeneous polynomial in 3 variables such that its partial derivatives have no common zeros in $\mathbb{R}^3 \setminus 0$. It is worth noticing that d is even if and only if all the curve components are two-sided, so that in the case of algebraic curves the degree parity introduced above and the parity of the algebraic degree d coincide.

Even if the curves are algebraic, there is no any restriction on the encoding tree as long as no condition on the curve is imposed. The situation is changing as soon as we fix the degree d of the curve. Then, already the number of connected components, and thus the number of the vertices in the encoding tree, is not arbitrary. As is known, the number of components of the curve is $\leq \frac{(d-1)(d-2)}{2} + 1$. Introduce, thus, the following notation which provides the sharp upper bound for the number of the vertices:

$$N_d = \begin{cases} (d-1)(d-2)/2 + 1, & \text{if } d \text{ is odd,} \\ (d-1)(d-2)/2 + 2, & \text{if } d \text{ is even.} \end{cases}$$

Starting from $d = 4$, not any tree with $\leq N_d$ vertices can be realized by a curve of degree d. Let I_d be the number of the trees which can be realized by curves of degree d. No direct formula or functional equation for these numbers is known;

moreover, their exact values are available only for $d \leq 7$. Very few is known even on the rate of growth of I_d.

As is shown in [5],

$$I_d \underset{e}{\asymp} \exp(d^2),$$

where $a_n \underset{e}{\asymp} b_n$ means that $\log a_n = O(\log b_n)$ and $\log b_n = O(\log a_n)$. On the other hand, due to Otter [6] (see also [3; Section 9.5]), one has the following exponential equivalence for the number T_n of rooted unlabeled trees with n vertices

$$T_n \underset{e}{\sim} C^n, \qquad C = 2.95576\ldots, \tag{1}$$

where the latter means that $\log T_n \sim n \log C$. This implies that

$$T_1 + \cdots + T_n \underset{e}{\sim} C^n,$$

hence,

$$I_d \leq C^{\frac{d^2}{2}+o(d^2)}. \tag{2}$$

The aim of the present note is to correct one erroneous remark from [5] and to show that the so-called Arnold inequalities [1] allow to reduce the constant C in the estimate (2). Namely, we prove that according to these inequalities

$$I_d \leq C_1^{\frac{d^2}{2}+o(d^2)}, \quad C_1 = 2.9193800\ldots \tag{3}$$

More precisely, $(\log C_1)d^2/2$ *is asymptotically equivalent to* $\log A_d$ *where* A_d *is the number of unlabeled trees with* $n \leq N_d$ *vertices not excluded by the Arnold inequalities.* According to [5], it implies that the Arnold inequalities exclude more arrangements of $\leq N_d$ closed simple circuits than any other known property of plane algebraic curves, including the consequences of the Bezout theorem.

Let us recall that the principal Arnold inequalities concern the curves of even degree $d = 2k$ exclusively. They state that

$$\text{even}^* \leq \frac{(k-1)(k-2)}{2} + 1, \quad \text{odd}^* \leq \frac{(k-1)(k-2)}{2}, \tag{4}$$

where even* is the number of internal vertices of odd distance from the root, and odd* is the number of internal vertices of even non-zero distance from the root. These inequalities imply the following lower bounds on the number l of leaves whatever is the parity of d:

$$l \geq n - 1 - \left[\frac{d-1}{2}\right]\left(\left[\frac{d-1}{2}\right] - 1\right) \tag{5}$$

where n is the total number of vertices. If d is even it is a straightforward consequence of (4) and if d is odd it follows from (5) for $d+1$. In particular, for the maximal value $n = N_d$ of n, the right hand side is approximately the half of n:

$$N_d - 1 - \left[\frac{d-1}{2}\right]\left(\left[\frac{d-1}{2}\right] - 1\right) \sim \frac{1}{2}N_d. \tag{6}$$

According to results of this note, it is the trees with $n = N_d$ and $l \sim \frac{1}{2}N_d$ which determine the asymptotical impact of Arnold bounds: *A_d has the same $\underset{e}{\asymp}$-rate of growth as the number of the trees with N_d vertices half of which are leaves.* In particular, the upper bound for I_{2k} deduced from the sole inequality (5) has the same $\underset{e}{\asymp}$-rate of growth as the upper bound which can be deduced from (4).

In fact, what is important in the coefficient 1/2 in (6) is that $1/2 > 0.438156\dots$ If the Arnold inequalities were not known but someone proved only that $l > 0.43N_d$, this fact would not reduce the constant C in (2) because the most of trees have about 43.8% leaves (see Appendix for details and references).

The note is organized as follows. The asymptotic growth of the number of the trees half of whose vertices are leaves is established in Section 1 in Theorem 7. The asymptotic impact of the Arnold inequalities is deduced from this theorem in Section 2: Theorem 9 takes into account only the bound (5) and Theorem 13 shows that (4) does not improve the rate. In Appendix we compare the result with the limiting distribution and show that the central limit theorem is not sufficient for our purpose: the range of values we treat is outside the range of a suitably good convergence.

1. On trees half of whose vertices are leaves.

1.1. Functional equation. Let us denote the number of rooted unlabeled trees with n vertices and m leaves by $a_{n,m}$ and consider the associated bi-variant generating function (a formal power series)

$$T(x,z) = \sum_{n,m} a_{n,m} x^n z^m = \sum_{n=1}^{\infty} a_n(z) x^n. \tag{7}$$

We get (see Figure 2)

$$T(x,z) = zx + zx^2 + (z+z^2)x^3 + (z+2z^2+z^3)x^4 + (z+4z^2+3z^3+z^4)x^5 + \\ (z+6z^2+8z^3+4z^4+z^5)x^6 + (z+9z^2+18z^3+14z^4+5z^5+z^6)x^7 + \dots$$

For technical reasons, we introduce also

$$\tilde{T}(x,z) = T(x,z) - zx + x = \sum_{n=1}^{\infty} \tilde{a}_n(z) x^n, \qquad \tilde{a}_n(z) = \begin{cases} 1, & n = 1 \\ a_n(z), & n > 1, \end{cases}$$

which is the generating function under the convention that the vertex of the one-vertex tree is not considered as a leaf.

Using Pólya enumeration theorem as it is done in [8] one can prove that $T(x,z)$ satisfies the (formal) functional equation

$$\tilde{T}(x,z) = T(x,z) - zx + x = x \exp\left(\sum_{k=1}^{\infty} \frac{T(x^k, z^k)}{k}\right). \tag{8}$$

The specialization $T(x) = T(x,1)$ is the classical generating function for the number of rooted unlabeled trees and substituting of $z = 1$ into (8) turns it into the classical Pólya equation, see [7].

It may be worth noticing that to prove (8), one can use as well the following bi-variant analog of the Cayley product formula for $T(x)$, cf. [4, formula 2.3.4.4-(3)],

$$\tilde{T}(x,z) = \frac{x}{\prod(1 - x^n z^m)^{a_{n,m}}}.$$

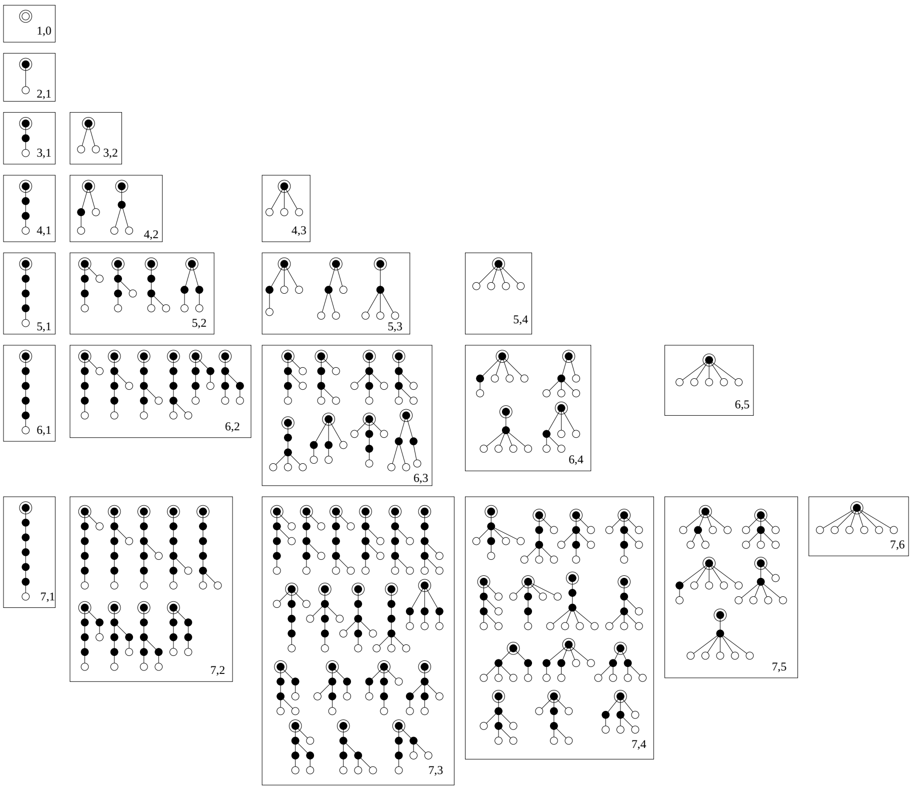

FIGURE 2. THREES WITH $n \le 7$ VERTICES

1.2. Recurrent relation. Taking the logarithmic derivatives of the both sides of (8), we get

$$\frac{\tilde{T}_x(x,z)}{\tilde{T}(x,z)} = \frac{\partial}{\partial x}\left(\log x + \sum_{k=1}^{\infty} \frac{T(x^k, z^k)}{k}\right) = \frac{1}{x} + \sum_{k=1}^{\infty} x^{k-1} T_x(x^k, z^k)$$

Multiplying the both sides by $x\,\tilde{T}(x,z)$ and subtracting $\tilde{T}(x,z)$, this gives

$$x\tilde{T}_x(x,z) - \tilde{T}(x,z) = \tilde{T}(x,z) \sum_{k=1}^{\infty} x^k T_x(x^k, z^k).$$

Hence,

$$\sum_{n=1}^{\infty} n\tilde{a}_{n+1}x^{n+1} = \sum_{p=1}^{\infty} \tilde{a}_p x^p \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} j a_j(z^k) x^{jk} = \sum_{n=1}^{\infty} x^{n+1} \sum_{p+jk=n+1} j a_j(z^k)\tilde{a}_p(z).$$

Thus, we obtain the recurrence relation (cf. [6] and [8])

$$n a_{n+1}(z) = n\tilde{a}_{n+1}(z) = \sum_{j=1}^{n} j \sum_{k=1}^{[n/j]} a_j(z^k)\tilde{a}_{n+1-jk}(z). \tag{9}$$

Together with the initial conditions $a_1(z) = z$, $\tilde{a}_1(z) = 1$, the relation (9) gives a rather fast way to compute $a_n(z)$.

1.3. Analytic properties of $T(x,z)$. If before we treated the generating functions as formal series, now we need to study their analytic behavior.

Let α be the radius of convergence of the power series $T(x)$. Using Polya's approach, see [7], i.e., resolving the equation $x\exp\left(1+\sum_{k=2}^{\infty}\frac{T(x^k)}{k}\right)=1$ (for instance, by Newton's method), one can compute α with any given precision. Indeed, any finite number of coefficients of the involved series can be computed using (9) and the number of terms to be summated, can be found from some rough estimate of α. Performing this computation, one gets

$$\alpha = 0.3383218568992076951961126257170170531 8\dots$$

This constant is sometimes called Otter constant because the first seven digits were computed in [6] (using the above approach from [7]).

Let us denote by D the domain of convergence of the series (7). Here, we follow the classical tradition and mean by the *domain of convergence* the interior of the set where the series is convergent. As is known, it coincides with the interior of the set of points $(x,z)\in\mathbb{C}^2$ such that $\sup_{n,m}|a_{n,m}x^nz^m|<\infty$. An important, also well known, consequence is that the logarithmic image

$$\log|D| = \{(\log|x|,\log|z|)\,:\,(x,z)\in D\}\subset\mathbb{R}^2$$

of any convergence domain is convex (in other words, the convergence domains are *logarithmically convex*).

Lemma 1. *There exists a continuous function $\zeta\mapsto r(\zeta),\mathbb{R}_{>0}=\{\zeta>0\}\to\mathbb{R}_{>0}$, such that $D=\{(x,z)\,:\,|x|<r(|z|)\}$. Moreover, $\alpha/\zeta\le r(\zeta)\le\alpha$ for $\zeta\ge 1$ and $r(\zeta)<\min\{1,\frac{1}{|\zeta|}\}$ for any ζ.*

The series $T(x,z)$ converges at each point $x=r(z), z>0$, of $\partial D\cap\mathbb{R}^2_{>0}$.

Proof. The existence statement and the nonsharp bounds follow from the logarithmic convexity of D combined with the cited above convergency properties of $T(x)=T(x,1)$ and with the fact that $D\subset\{|xz|\le 1\}$; in its turn, this inclusion follows from $a_{n,m}\ge 1$ for any $n>m$. The strict inequality $r(z)<\frac{1}{|z|}$ is a consequence of the convergence of $T(x,z)$ at the boundary points. To prove this convergence it sufficient to notice that

$$T(x,z)=xz-x+xe^{T(x,z)+\cdots}>xz-x+xe^{T(x,z)}$$

for $0<x<r(z), z>0$; it implies the boundedness of T on the interval $x\in[0,r(z)[$ and, by Abel theorem, its convergence at $x=r(z)$. $\square$

Lemma 2. *The transformations $(x,z)\mapsto(x^k,z^k), k\ge 2$, map D into itself. For any point $(x,z), z\ne 0$, in the closure of D the series*

$$h(x,z)=\sum_{k=2}^{\infty}\frac{T(x^k,z^k)}{k}$$

is absolutely convergent and defines a function holomorphic at such a point.

Proof. The invariance property follows from the logarithmic convexity and the bounds on $r(z)$ given by Lemma 1. In addition, due to this Lemma, for all

(x, z) in a small neighborhood of any point in the closure of D we have bounds $|x^k| \leq a^k, |z^k| \leq b^k$ with $a < 1, ab < 1$ whatever is $k \geq 1$. These bounds provide a bounded convergence of the series:

$$\sum_{k\geq 2}\sum_{n,m} \frac{|a_{n,m}x^{nk}z^{mk}|}{k} \leq \sum_{n,m}\sum_{k\geq 2} \frac{a_{n,m}a^{nk}b^{mk}}{k} \leq \lambda \sum_{n,m} a_{n,m}a^{2n}b^{2m} = \lambda T(a^2, b^2). \quad \square$$

In what follows we study the boundary values $a(z) = T(r(z), z), z > 0$ of T and use an auxiliary function

$$F(x, y, z) = z - 1 + e^{y+h(x,z)} - \frac{y}{x}.$$

By (8), we have $F(x, T(x, z), z) = 0$ at any point of the closure of D with $x \neq 0, z \neq 0$. In particular, the real curve $x = r(z), z > 0$, satisfies the equation

$$F(x, a(z), z) = 0.$$

Lemma 3. *The function $r(\zeta)$ is analytic. The function $F(x, y, z)$ is analytic near the real curve $x = r(z)$, $z > 0$. We have*

$$F_y(r(z), a(z), z) = 0, \tag{10}$$

$$a(z) = 1 + r(z)(z - 1). \tag{11}$$

Proof. The analyticity of F follows from Lemma 2, and then all the other statements, except the relation (11), follow from the implicit function theorem.

Let us show that $a(z) = 1 + r(z)(z - 1)$. By the definition of F, we have $F_y = e^{y+h(x,z)} - \frac{1}{x}$. Hence, for $x = r(z)$ and $y = a(z)$ we have

$$0 = F_y = e^{y+h(x,z)} - \frac{1}{x} \qquad \text{and} \qquad 0 = F = z - 1 + e^{y+h(x,z)} - \frac{y}{x}.$$

Thus, $\frac{y}{x} = \frac{1}{x} + z - 1$ and $y = 1 + x(z - 1)$. $\square$

Due to Lemma 3, the function $x = r(z)$ can be found by resolving the equation

$$x \exp(1 + (z - 1)x + h(x, z)) = 1$$

This allows one to compute $r(z)$ with any given precision.

Let us define

$$a_n^+(z) = \sum_{m>n/2} a_{n,m}z^m, \quad a_n^-(z) = \sum_{m\leq n/2} a_{n,m}z^m, \quad T_\pm(x, z) = \sum_{n=1}^{\infty} a_n^\pm(z)x^n;$$
$$\hat{T}(x, z) = \hat{T}(xz^{-1/2}, z), \quad \hat{T}_\pm(x, z) = \hat{T}_\pm(xz^{-1/2}, z), \quad \hat{r}(\zeta) = r(\zeta)\sqrt{\zeta}.$$

and denote by $\hat{D}$ and $\hat{D}_\pm$ the domain of convergence of $\hat{T}$ and $\hat{T}_\pm$, respectively. It is clear that $\hat{D} = \{(x, z) \,:\, |x| < \hat{r}(|z|)\}$.

Lemma 4. *The function $\hat{r}(\zeta)$ has a single critical point, this point is a point of maximum.*

Proof. The logarithmic map $(x, z) \mapsto (\log|x|, \log|z|)$ transforms $xz^{-\frac{1}{2}}$ in a linear function. Therefore, due to the convexity of $\log|D|$, the critical points of $\hat{r}(\zeta)$ form a convex set. If it is not reduced to a single point, then, since r is real analytic, $r(\zeta) = c\zeta^{-\frac{1}{2}}, c > 0$, which contradicts to the bounds from Lemma 1.

It is a point of maximum, since the domains of convergence are Reinhardt domains, i.e., $(x, z) \in D$ as soon as there exists $(x_0, z_0) \in D$ with $|x| < |x_0|, |z| < |z_0|$. $\square$

Denote by $z_0 \in \mathbb{R}_+$ the point where the maximum of $\hat{r}(\zeta)$ is attained and put $x_0 = \hat{r}(z_0)$.

Proposition 5. $\hat{D}_\pm = \{(x, z) : |x| < \hat{r}_\pm(|z|)\}$, *where*

$$\hat{r}_-(\zeta) = \max_{\omega \le \zeta} \hat{r}(\omega) = \begin{cases} \hat{r}(\zeta), & \zeta \le z_0, \\ x_0, & \zeta \ge z_0, \end{cases} \quad \text{and} \quad \hat{r}_+(\zeta) = \max_{\omega \ge \zeta} \hat{r}(\omega) = \begin{cases} \hat{r}(\zeta), & \zeta \ge z_0, \\ x_0, & \zeta \le z_0. \end{cases}$$

Proof. For a point $p = (u_0, v_0) \in \mathbb{R}^2$, let us denote $\mathbb{R}^2_{+-}(p) = \{(u, v) \,|\, u \le u_0, v \ge v_0\}$ and $\mathbb{R}^2_{++}(p) = \{(u, v) \,|\, u \le u_0, v \le v_0\}$. The result follows from the following properties:

(a) $\log|\hat{D}|$ and $\log|\hat{D}_\pm|$ are convex;
(b) If $p \in \log|\hat{D}_\pm|$ then $\mathbb{R}^2_{+\pm}(p) \subset \log|\hat{D}_\pm|$;
(c) $\log|\hat{D}| = \log|\hat{D}_+| \cap \log|\hat{D}_-|$. $\square$

1.4. Rate of growth.

Theorem 6.

$$\sum_{m>n/2} a_{n,m} \underset{e}{\sim} C_1^n, \qquad \text{where} \quad C_1 = \frac{1}{x_0} = 2.919380017448416911265032583985\dots$$

Proof. The coefficients $a_n^+(1) = \sum_{m>n/2} a_{n,m}$ of the power series $\hat{T}_+(x, 1)$ satisfy the following relation

$$\log a^+_{n+m+2}(1) \ge \log a^+_n(1) + \log a^+_m(1) - \log 2$$

(to prove this relation it is sufficient to plant two trees over a new root and to add a leaf growing from the root). Hence, the sequence $n^{-1} \log a_n^+(1)$ has a limit and, by the Cauchy rule,

$$\sum_{m>n/2} a_{n,m} = a_n^+(1) \underset{e}{\sim} \hat{r}_+(1)^{-n}. \tag{12}$$

To compute $\hat{r}_+(1)$, we must find z_0. We compute it as the root of the equation $\hat{r}'(z) = 0$ (the root is unique by the convexity of $\log D$). To find it by Newton's method, we need $\hat{r}'(z)$ and $\hat{r}''(z)$. They can be found as follows. Derivating the identity $F(r(z), a(z), z) = 0$ and using (10), we get

$$F_x(r(z), a(z), z) r' + F_z(r(z), a(z), z) = 0. \tag{13}$$

Derivating again, we see that at points $(r(z), a(z), z)$ one has

$$F_{xx}r'^2 + F_{xy}r'a' + 2F_{xz}x' + F_{yz}a' + F_{zz} + F_x r'' = 0. \tag{14}$$

Note that a' can be found from (11).

The partial derivatives of F at a point $(r(z), a(z), z)$ are

$$F_x = (h_x/r) + (a/r^2), \qquad F_y = 0, \quad F_z = 1 + (h_z/r),$$

$$F_{xx} = (h_{xx}+h_x^2)/r - 2(a/r^3), \qquad F_{xy} = (h_x/r)+(1/r^2), \qquad F_{xz} = (h_{xz}+h_xh_z)/r,$$

$$F_{yz} = h_z/r, \qquad F_{zz} = (h_{zz} + h_z^2)/r$$

Solving the equation $\hat{r}'(z) = 0$ by Newton's method, we find

$$z_0 = 1.4849173957741380958 7489\ldots$$

and

$$x_0 = \hat{r}(z_0) = 0.34253848215143138449599 19944869\ldots$$

Since $z_0 > 1$, we have $\hat{r}_+(1) = \hat{r}(z_0) = x_0$. Now, the desired asymptotic relation follows from (12) and

$$C_1 = 1/x_0 = 2.91938001744841691126503258 3985\ldots \qquad \square$$

Theorem 7. *There is a continuous function $\lambda \mapsto C(\lambda)$, $\mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$, such that*

$$\sum_{m>\lambda n} a_{n,m} \underset{e}{\sim} C(\lambda)^n \qquad \text{for any } \lambda \geq 0.$$

For each $\lambda > \frac{1}{2}$ one has $C(\lambda) < C(\frac{1}{2}) = C_1$.

Proof. Let $z_{0,\lambda}$ be the critical point of $r(\zeta)\zeta^\lambda$. By the same arguments as in the proof of Proposition 5 and Theorem 6,

$$\sum_{m>\lambda n} a_{n,m} \underset{e}{\sim} \hat{r}_{+,\lambda}(1)^{-n},$$

where $\hat{r}_{+,\lambda}(1)$ is equal to $r(z_{0,\lambda})z_{0,\lambda}^\lambda$ if $1 < z_{0,\lambda}$ and to $r(1)$ otherwise. Due to the logarithmic convexity of D,

$$z_{0,\lambda} > z_0 \quad \text{and} \quad r(z_{0,\lambda})z_{0,\lambda}^\lambda > r(z_0)z_0^\lambda > r(z_0)z_0^{\frac{1}{2}},$$

if $\lambda > \frac{1}{2}$. $\square$

2. On the impact of Arnold inequalities.

2.1. Impact of the bound on the number of nonempty ovals. Consider first the case of curves of degree d with $(d-1)(d-2)/2+1$ connected components and denote by L_d the number of the trees which satisfy the Arnold bound (5). Namely, L_d is the number of rooted unlabeled trees with $n = N_d$ vertices and $\geq M_d$ leaves where $M_d = N_d - 1 - [\frac{d-1}{2}]([\frac{d-1}{2}]-1)$. Recall that $N_d \sim \frac{d^2}{2}$ (see also (6)).

Proposition 8.

$$L_d \underset{e}{\sim} C_1^{\frac{d^2}{2}}.$$

Proof. We apply Theorem 7. Since $C(\lambda)$ is continuous at $\lambda = \frac{1}{2}$, we find for any $\epsilon > 0$ such $\delta > 0$ that for any sufficiently big n it holds

$$(C_1+\epsilon)^{(1+\epsilon)n} \geq \sum_{m>(\frac{1}{2}-\delta)n} a_{n,m} \quad \text{and} \quad \sum_{m>(\frac{1}{2}+\delta)n} a_{n,m} \geq (C_1-\epsilon)^{(1-\epsilon)n}.$$

It remains to put $n = N_d \sim \frac{d^2}{2}$ and to note that for any sufficiently big d

$$\sum_{m>(\frac{1}{2}-\delta)n} a_{n,m} \geq L_d \geq \sum_{m>(\frac{1}{2}+\delta)n} a_{n,m}. \quad \square$$

Now, consider the general case and denote, in accordance with the Arnold bound on the number of empty ovals, by L'_d the number of rooted unlabeled trees with $n \leq N_d$ vertices and $\geq n - [\frac{d-1}{2}]([\frac{d-1}{2}]-1)$ leaves.

Theorem 9.

$$L'_d \underset{e}{\sim} C_1^{\frac{d^2}{2}}.$$

Proof. In view of (1) and Proposition 8, it is sufficient to prove that $L'_d \leq (\hat{k}^2 - \hat{k})T_{\hat{k}^2-\hat{k}} + k^2 L_d$ where $\hat{k} = [\frac{d-1}{2}]$ and $k = [\frac{d}{2}]$. Clearly, the first term bounds from above the total number of trees with $n \leq \hat{k}^2 - \hat{k}$ vertices. In the range $\hat{k}^2 - \hat{k} < n \leq N_d$ the number of the trees excluded by the Arnold bound (5) is increasing, from 0 to L_d, when n grows, since $a_{n,m} \leq a_{n+1,m+1}$ (to prove such an inequality it is sufficient to add a leaf to a branch with a maximal number of leaves). The coefficient k^2 before L_d is due to

$$N_d - 1 - [\frac{d-1}{2}]([\frac{d-1}{2}]-1) = k^2. \quad \square$$

2.2. Auxiliary lemmas.

Let v be a vertex of a tree t. A *branch of t at v* is a connected component of the graph obtained from t by removing v and the (open) edges adjacent to v.

Lemma 10. *Let t be a tree with N vertices. Then there exists a vertex v such that any branch of t at v has at most $N/2$ vertices.*

Proof. Suppose that any vertex has a branch with more than $N/2$ vertices. Choose any vertex v_1 and define the sequence of vertices $v_1, v_2, \dots$ as follows. Assume that v_i is already defined. Let t_i be the branch of t at v_i which has more than $N/2$ vertices. Then v_{i+1} is defined as the vertex of t_i which is nearest to v_i. Moving from v_1 to v_2, then from v_2 to v_3 and so on, we can never turn back. Indeed, if v_{i+1} coincides with v_{i-1} then removing from t the (open) edge connecting v_i with v_{i+1} we would obtain two subtrees of t each having more than $N/2$ vertices. Since t has no loops, this means that our sequence has no repeatings. Contradiction. $\square$

Lemma 11. *Let $c_1 \geq \dots \geq c_r \geq 0$ and $|c| \leq c_1 + \dots + c_r$. Then there exist $\varepsilon_1, \dots, \varepsilon_r \in \{\pm 1\}$ such that $|(\varepsilon_2 c_2 + \dots + \varepsilon_r c_r) - c| \leq c_1$.*

Proof. Set $\varepsilon_{k+1} = \operatorname{sign}_+(c - (\varepsilon_2 c_2 + \dots + \varepsilon_k c_k))$ where $\operatorname{sign}_+(x) = 1$ for $x \geq 0$ and $\operatorname{sign}_+(x) = -1$ for $x < 0$. This means that we walk along the real axis starting from the origin so that the absolute values of the steps are successively $c_2, c_3, \dots$ and each step is directed towards the point c. Then $c' = \varepsilon_2 c_2 + \dots + \varepsilon_r c_r$ is the final point of our walk. It is easy to see that $|c' - c| \leq c_1$. $\square$

In accordance with the terminology coming from the geometry of plane curves, let us say that a vertex of a rooted tree t is *even* (resp. *odd*) if the minimal path relating it to the root consists of an odd (resp. even) number of edges. Let us denote by $p(t)$ (resp. $n(t)$) the number of even (resp. odd) vertices, including the root, of t and put $\chi(t) = p(t) - n(t)$.

For example, the root is an odd vertex, the vertices connected to the root by an edge are even etc. Note, that when we change the root, $|\chi(t)|$ does not change.

We say that a rooted tree t' is obtained from a rooted tree t by *contracting an edge* if t' is obtained from t by replacing some edge with a single vertex v (see Figure 3). If one of the ends of the edge which we contracted was the root of t, then v is declared the root of t'. This operation reduces the number of vertices and the edges by one. The operation of *inserting an edge at v* is to be thought of as an inverse operation. When one of the ends of the inserted edge is a leaf, this is called the *attachment of an edge.*

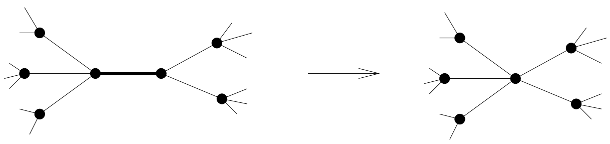

FIGURE 3. EDGE CONTRACTING

Lemma 12. *Let t_0 be a rooted tree with N vertices and let c be any integer such that $|c| \leq |\chi(t_0)|$. Then there exists a sequence of rooted trees $t_1, \dots, t_k$ such that*

(1) *$\chi(t_k) = c$;*
(2) *t_{i-1}, $i = 1, \dots, k$, is obtained from t_i by contracting an edge;*
(3) *$k \leq 3 + 3 \log_2 N$.*

Proof. Apply the induction by N. The case $N = 1$ is trivial. Assume that the statement is true for any tree which has less than $N > 1$ vertices. By Lemma 10, there exists a vertex v such that any branch of t_0 has at most $N/2$ vertices. Let us denote the branches of t at v by $b_1, \dots, b_r$. We choose the root of each branch at the vertex nearest to v. Let $c_i = |\chi(b_i)|$ and $\delta_i = \operatorname{sign}\chi(b_i)$. Let us number the branches so that $c_1 \ge c_2 \ge \dots \ge c_r$. By Lemma 11, there exist $\varepsilon_2, \dots, \varepsilon_r \in \{\pm 1\}$ such that $|c' - c| \le c_1$ where $c' = \varepsilon_2 c_2 + \dots + \varepsilon_r c_r$. By the induction hypothesis, we can insert $\le 3 + 3\log(N/2) = 3\log N$ edges to b_1 so that $c_1^* = \chi(b_1^*) = c - c'$ for the resulting tree b_1^*. Let t_0^* be the tree obtained from t_0 by replacing b_1 with b_1^*.

Let t_1^* be obtained from t_0^* by inserting an edge e at v so that b_1^* and the branches $b_i, i \ge 1$, with $\varepsilon_i\delta_i = \operatorname{sign}(c - c')$ are on one side of e and the branches b_i with $\varepsilon_i\delta_i = -\operatorname{sign}(c - c')$ are on the other side. Then we have $|\chi(t_1^*)| = |c_1^* + c'| = |c|$. Now, we may return to counting $|\chi|$ with respect to the initial root of t_0 and respective roots of $t_i, i \ge 1$. If $\chi(t_1^*) = -c$, we attach an edge to the root, choose the obtained leaf as the new root and then attach an edge to the new root. $\square$

2.3. Impact of the bounds on the number of even and odd nonempty ovals.

Let us recall that A_d denotes the number of rooted unlabeled trees with $n \le N_d$ vertices which satisfy the Arnold bounds (4).

Theorem 13.

$$A_d \underset{e}{\sim} C_1^{\frac{d^2}{2}}.$$

Proof. If a tree with $n \le N_d$ vertices satisfies the weak Arnold bound (5), we apply to it, removing its leaves, Lemma 12 with $c = 0$, and then put the leaves back, getting thus a tree with $n + 3[\log_2 n] + 3 \le N_d + 3[\log_2 N_d] + 3 \le N_{d+6}$ vertices which satisfies the stronger Arnold bounds (4). Therefore,

$$\frac{L'_d}{A_{d+6}} \le \sum_{n=1}^{N_d} \binom{n + 3[\log_2 n] + 3}{3[\log_2 n] + 3} \le N_d \binom{N_d + 3[\log_2 N_d] + 3}{3[\log_2 N_d] + 3} = e^{o(N_d)}$$

and the theorem follows now from Theorem 9 and $A_d \le L'_d$. $\square$

Appendix. Limit distribution

Let us consider $a_{n,m}/a_n(1)$ as a probability distribution of a random variable X_n, i.e. $P(X_n = m) = a_{n,m}/a_n(1)$. As is known, see f.e. [2], the following central limit theorem holds: this random sequence X_n, once normalized, tends to a normal distribution:

$$P(a < \frac{X_n - \mu n}{\sigma\sqrt{n}} < b) \to \frac{1}{2\pi}\int_a^b e^{-\frac{x^2}{2}}\,dx.$$

where

$$\mu = -r'(1)/\alpha = 0.43815623566437466396849216386287978370 55\ldots$$

and

$$\sigma^2 = \frac{r'(1)^2}{\alpha^2} - \frac{r'(1) + r''(1)}{\alpha} = 0.150044811672846981980699640444640111071\ldots$$

In particular, this means that approximately 43.8% of vertices of a big random tree are leaves. The fact that the mean value of the number of leaves is $\sim \mu n, \mu =$

0.438156235664... was established by Robinson and Schwenk [8] by the Polya-Otter method, and its extension to the other moments was given by Schwenk in [9].

In view of the above limit theorem, it is natural to replace $a_{n,m}$ by its approximation by the normal distribution

$$a^*_{n,m} = \frac{a_n(1)}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(m-\mu n)^2}{2\sigma^2 n}\right).$$

Then, we get

$$\sum_{m>n/2} a^*_{n,m} \underset{e}{\sim} \alpha^{-n} \exp\left(-\frac{(1/2-\mu)^2 n}{2\sigma^2}\right) = C_2^n$$

where

$$C_2 = \alpha^{-1} \exp\left(-\frac{(1/2-\mu)^2}{2\sigma^2}\right) = 2.918333013459557401497869878213291811 93\ldots$$

We see that C_2 differs from C_1 in the fourth digit. This is not a contradiction with the central limit theorem because this just means that the convergence to the normal distribution is not good far from the center. It shows that the central limit theorem is not sufficient for a search of the rate of growth of $\sum_{m>n/2} a_{n,m}$.

To conclude, let us notice that the constants $r'(1)$ and $r''(1)$ (needed to find μ and σ^2) can be computed much faster than the constants z_0 and x_0 from Section 2 because the double summation over n, m may be replaced with the single summation by use of the following recurrent formulas for the coefficients of the series $T_z(x,1)$ and $T_{zz}(x,1)$. Similarly to (9), one can obtain

$$a'_{n+1}(1) = \sum_{j=1}^{n} a'_j(1) \sum_{k=1}^{[n/j]} a_{n+1-kj}(1)$$

$$\begin{aligned} a''_{n+1}(1) = \sum_{j=1}^{n} \Big\{ & a'_j(1)\Big(\sum_{k=1}^{[(n-1)/j]} a'_{n+1-kj}(1)\Big) \\ & + a'_j(1)\Big(\sum_{k=1}^{[n/j]} (k-1)a_{n+1-kj}(1)\Big) + a''_j(1)\Big(\sum_{k=1}^{[n/k]} k a_{n+1-kj}(1)\Big)\Big\} \end{aligned}$$

References

1. V.I. Arnold, *On arrangement of ovals of real plane algebraic curves, the involutions of four-dimensional smooth manifolds, and the arithmetic of integer-valued quadratic forms*, Funkcional. Anal. i Priložen **5** (1971), 1–9 (Russian); English transl., Funct. Anal. and Appl. **5** (1971), 169–176.
2. M. Drmota, B. Gittenberger, *The distribution of nodes of given degree in random trees*, J. of Graph Theory **31** (1999), 227-253.
3. F. Harary, E. Palmer, *Graphical enumeration*, Academic Press, New York, London, 1973.
4. D. Knuth, *The art of computer programming*, vol. 1, Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont, 1968.

5. S.Yu. Orevkov, V.M. Kharlamov, *Asymptotic growth of the number of classes of real plane algebraic curves as the degree grows*, Zapiski Nauchn. Semin. POMI **266** (2000), 218–233 (Russian); English transl., J. of Math. Sciences **113** (2003), 666–674.
6. R. Otter, *The number of trees*, Ann. Math. **49** (1948), 583–599.
7. G.Pólya, *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen*, Acta Math. **49** (1937), 145–254.
8. R.W.Robinson, A.J.Schwenk, *The distribution of degrees in a large random tree*, Discrete Math. **12** (1975), 359–372.
9. A.J. Schwenk, *An asymptotic evaluation of the cycle index of a symmetric group*, Discrete Math. **18** (1977), 71–78.

First author: Université Louis Pasteur et IRMA (CNRS), 7 rue René Descartes 67084 Strasbourg Cedex, France

E-mail address: kharlam@math.u-strasbg.fr

Second author: Laboratoire Emile Picard, UFR MIG, Univ. Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France

E-mail address: orevkov@picard.ups-tlse.fr