

Real algebraic curves, the moment map and amoebas

By G. MIKHALKIN*

Abstract

In this paper we prove the topological uniqueness of maximal arrangements of a real plane algebraic curve with respect to three lines. More generally, we prove the topological uniqueness of a maximally arranged algebraic curve on a real toric surface. We use the moment map as a tool for studying the topology of real algebraic curves and their complexifications.

1. Introduction and statement of results

1.1. *M-curves in the plane.* An algebraic curve $\mathbb{R}\bar{A} \subset \mathbb{R}P^2$ is the zero set of a polynomial p of degree d . (We reserve the notation $\mathbb{R}A$ for the corresponding curve in $(\mathbb{R} - 0)^2$.) Suppose that $\mathbb{R}\bar{A}$ is nonsingular. Then it is homeomorphic to a disjoint union of circles. If d is even then each component of $\mathbb{R}\bar{A}$ bounds a disk in $\mathbb{R}P^2$; such a component is called an *oval*. If d is odd then all the components but one are ovals and the remaining one is a one-sided circle in $\mathbb{R}P^2$. Harnack's inequality [5] states that the number of components is not greater than $\frac{(d-1)(d-2)}{2} + 1$. If it is equal to $\frac{(d-1)(d-2)}{2} + 1$ then $\mathbb{R}\bar{A}$ is called an *M-curve*. Let l_1, \dots, l_n be lines in general position in $\mathbb{R}P^2$ (i.e. no three lines pass through the same point).

Definition 1. We say that $\mathbb{R}\bar{A}$ is in *maximal position* with respect to a collection of n lines l_1, \dots, l_n if $\mathbb{R}\bar{A}$ is an M-curve and there exist n disjoint arcs $a_1, \dots, a_n \subset \mathbb{R}\bar{A}$ such that a_j intersect l_j in d points and all the arcs belong to the same component of $\mathbb{R}\bar{A}$; see Figure 1.

Remark 1. The arcs from Definition 1 were called *bases of rank 1* in [2]. Brusotti used them to generalize the constructions of Harnack [5] and Hilbert [6] and produce a larger variety of M-curves for all d . A starting curve for Brusotti's construction is an M-curve with 2 disjoint bases (possibly of higher

*Research was partially supported by the NSF (DMS#9801726).

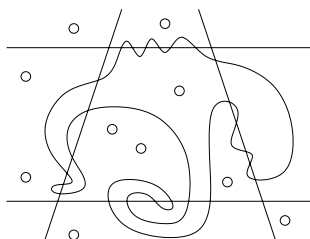


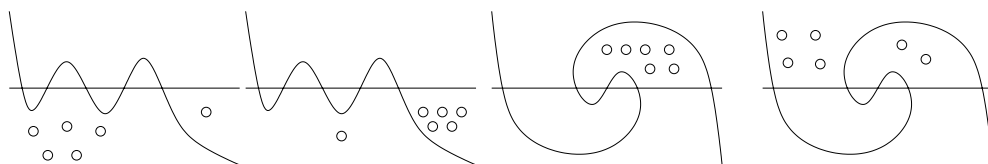
Figure 1. Maximal position with respect to lines.

rank). The constructions of Harnack and Hilbert are special cases where one takes a line or an ellipse (respectively) for the starting curve. Note that both the line and the ellipse are M-curves with arbitrarily many disjoint bases of rank 1.

We call the topological type of $(\mathbb{R}P^2; \mathbb{R}\bar{A}, l_1 \cup \dots \cup l_n)$, where $\mathbb{R}\bar{A}$ is a curve of degree d in maximal position with respect to $l_1 \cup \dots \cup l_n$, a *maximal topological type*. What are maximal topological types for given n and d ?

For $n = 0$ this problem is a part of Hilbert's 16th problem [7]. It is an open question. There are powerful theorems known which show that the topological type of an M-curve is very restricted ([12], [1], [14] et al.). But on the other hand there is a large variety of M-curves in different topological types. The complete answer is known only for $d \leq 7$; see [15].

For $n = 1$ this asks for topological classification of affine M-curves. Indeed, the maximality condition from Definition 1 for $n = 1$ just states that the affine curve $\mathbb{R}\bar{A} - l_1 \subset \mathbb{R}^2 = \mathbb{R}P^2 - l_1$ has $\frac{(d-1)(d-2)}{2} + d$ components, the maximal possible value for a curve of degree d . The affine M-curves of degree 5 were classified in [13]; see Figure 2. The question is open for $d > 5$.

Figure 2. Maximal topological types for $n = 1$, $d = 5$.

For $n = 2$ there are several constructions of M-curves of the same degree d but of different topological types found by Brusotti [2] with $d \geq 4$. Figure 3 pictures all the types for $d = 4$. For $d > 4$ the question is open.

The following two theorems answer this question for all d if $n \geq 3$. For $n = 3$ we can assume that l_1, l_2, l_3 are coordinate lines in $\mathbb{R}P^2$ (the x -axis, the y -axis and the infinite line).

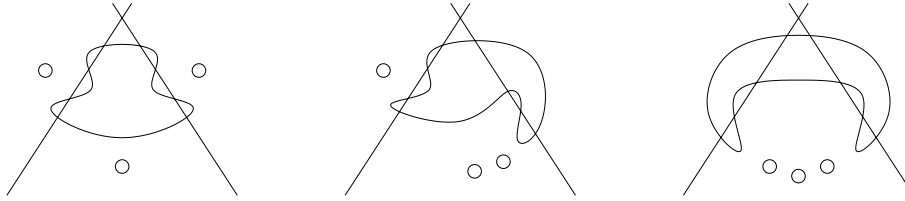


Figure 3. Maximal topological types for $n = 2, d = 4$.

THEOREM 1. *The maximal topological type is unique for $n = 3$ and any d . If $\mathbb{R}\bar{A} \subset \mathbb{R}P^2$ is a curve of degree d in maximal position with respect to the coordinate axes then the topological type of $(\mathbb{R}P^2; \mathbb{R}\bar{A}, l_1 \cup l_2 \cup l_3)$ depends only on d and is pictured in Figure 4 for even d or Figure 5 for odd d .*

The M-curves pictured in Figure 4 and Figure 5 were constructed by Harnack [5]. In fact, they were the first examples of M-curves in $\mathbb{R}P^2$ for arbitrary d .

THEOREM 2. *There are no maximal topological types for $n > 3$ and $d \geq 3$.*

In particular, this takes care of the case $n > 4$ where the choice of the lines $l_1, \dots, l_n \subset \mathbb{R}P^2$ is not unique so that the answer might presumably depend on it.

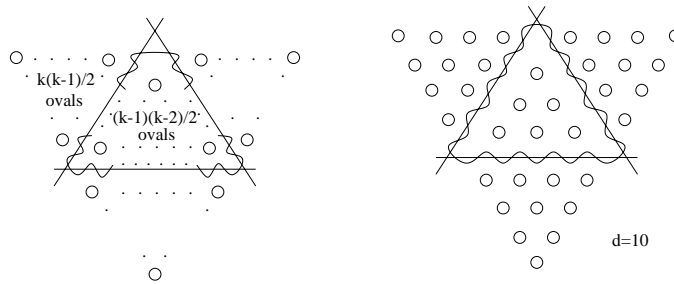


Figure 4. The maximal type for $n = 3, d = 2k$; e.g. $d = 10$.

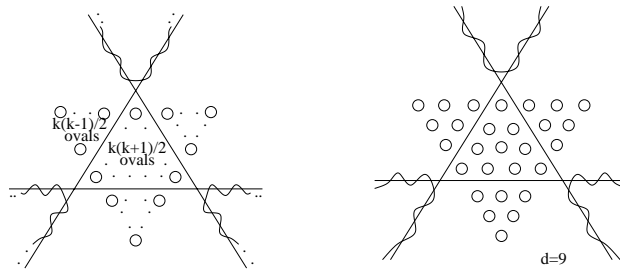


Figure 5. The maximal type for $n = 3, d = 2k + 1$; e.g. $d = 9$.

Remark 2. For every d there exists an algebraic curve of degree d in maximal position with respect to three lines and invariant with respect to an S_3 group of symmetries of these three lines. Such a curve can be constructed by patchworking; see the appendix. Note that the curves in Figure 4 intersect each of the lines in an infinite point (the topological arrangements pictured there also admit an S_3 -symmetry).

Remark 3. Let $d = 3$. An arc of a real cubic curve is a base of rank 1 if and only if it contains an inflection point. Thus, Theorem 2 may be viewed as a generalization of the fact that a plane cubic curve cannot have more than three real inflection points.

Remark 4. The condition that $\mathbb{R}\bar{A}$ is an M-curve is essential for Theorem 2. There exists a quartic with three ovals and with four bases of rank 1 on the same oval. The condition that the bases of rank 1 be on the same component is also essential. There exists an M-quartic with four bases of rank 1 on different ovals. Such quartics may be obtained by perturbing a union of four lines; see Figure 6. The first quartic consists of three ovals and two of them intersect the infinite line. The second quartic consists of four ovals, none intersects the infinite line.

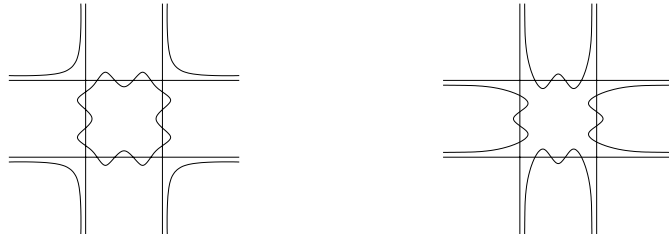


Figure 6. Quartics with four bases.

1.2. *M-curves in toric surfaces.* A convex polygon $\Delta \subset \mathbb{R}^2$ with vertices in \mathbb{Z}^2 defines a compactification of $(\mathbb{C} - 0)^2$ to an algebraic surface T called the toric surface; see e.g. [4]. The action of the torus $S^1 \times S^1$ on $(\mathbb{C} - 0)^2$ by $(\alpha, \beta) \times (x, y) = (\alpha x, \beta y)$, where we view each S^1 as the unit circle in \mathbb{C} , extends to the action of $S^1 \times S^1$ on T . The moment map $\mu_T : T \rightarrow \Delta$ of this action exhibits T as a singular Lagrangian fibration over Δ . The projection μ_T takes the quotient by the action of the group $S^1 \times S^1$. The closure $\mathbb{R}T$ of $(\mathbb{R} - 0)^2$ in T is the real toric surface and T is its complexification. Over the interior of Δ the fiber of μ_T is the full torus $S^1 \times S^1$. Over the sides of Δ the fiber is the quotient of the torus by the circle subgroup corresponding to the (rational) slope of the side. Over the vertices of Δ the fiber is one point. The inverse images of the sides of Δ are rational holomorphic curves in T ; we call them *the axes* of T .

Example. Let Δ be the triangle with vertices $(d, 0)$, $(0, d)$ and $(0, 0)$. For any d , the corresponding toric surfaces are the real projective plane $\mathbb{R}P^2$ and its complexification $\mathbb{C}P^2$ (different values of d correspond to different multiples of the Kähler form). The inverse images of the sides $[(0, 0), (d, 0)]$, $[(0, 0), (0, d)]$ and $[(d, 0), (0, d)]$ are the x -axis, the y -axis and the infinite line respectively.

Let $\mathbb{R}\bar{A} \subset \mathbb{R}T$ be an algebraic curve and $\bar{A} \subset T$ be its complexification. It is given by a real polynomial p in two variables. Recall that the Newton polygon Δ of $p = \sum a_{j,k}x^jy^k$ is the convex hull of $\{(j, k) \mid a_{(j,k)} \neq 0\}$ in \mathbb{R}^2 . If \bar{A} does not pass through the intersection of the axes of T but intersects every axis, then T is necessarily the toric surface corresponding to the Newton polygon Δ . If \bar{A} is nonsingular then the genus g of \bar{A} is equal to the number of lattice points in the interior of Δ ; see [10]. We call $\mathbb{R}\bar{A}$ an *M-curve* if the number of components of $\mathbb{R}\bar{A}$ is equal to $g + 1$; by Harnack's inequality this is the largest possible number. Let l_1, \dots, l_n be the axes of T in the order corresponding to the order of the sides of Δ . Let d_j be the integer length of l_j (one plus the number of integer points inside l_j). Note that d_j is the degree of the restriction of p to l_j and, therefore, the number of intersection points of l_j and $\mathbb{R}\bar{A}$ is no more than d_j .

Definition 2. We say that $\mathbb{R}\bar{A}$ is in *maximal position* in $\mathbb{R}T$ if $\mathbb{R}\bar{A}$ is an M-curve and there exist n disjoint arcs $c_1, \dots, c_n \subset \mathbb{R}\bar{A}$ such that the c_j intersect l_j in d_j points and all the arcs belong to the same component of $\mathbb{R}\bar{A}$. We say that $\mathbb{R}\bar{A}$ is in *cyclically maximal position* in $\mathbb{R}T$ if, in addition, the order of arcs c_j agrees with the cyclic order on the component of $\mathbb{R}\bar{A}$.

Remark 5. Obviously, if $n \leq 3$ then a curve in maximal position in $\mathbb{R}T$ is automatically in cyclically maximal position in $\mathbb{R}T$. The same is true if $n > 3$ and all d_j are even. See Example 1 in Section 4 for a curve in maximal but not in cyclically maximal position in $\mathbb{R}T$.

The following theorem is a generalization of Theorem 1.

THEOREM 3. *If $\mathbb{R}\bar{A}$ is in cyclically maximal position in $\mathbb{R}T$ then the topological type of $(\mathbb{R}T; \mathbb{R}\bar{A}, l_1 \cup \dots \cup l_n)$ depends only on Δ .*

The maximal topological type for Δ can be reconstructed from Figure 12 and Lemma 11.

Remark 6. The maximality definition and Theorem 3 can be restated in terms of curves in $(\mathbb{R} - 0)^2$.

Curves in maximal position in $\mathbb{R}T$ exist for any Δ and the corresponding toric surface T ; see Corollary A4 of the appendix.

Remark 7. For some shapes of Δ the surface T has singularities at the intersection of axes but (by the maximal position assumption) $\mathbb{R}\bar{A}$ does not pass through them.

Remark 8. Note that different polygons Δ with parallel sides correspond to different homology classes of \bar{A} in the same T .

2. Amoebas

In this section we introduce the tools for proving the main results. They are based on the following notion of *amoeba* defined by Gelfand et al. in [4].

Definition 3. Let $p : \mathbb{C}^m \rightarrow \mathbb{C}$ be a polynomial. Its amoeba is the image $\mu(A) \subset \mathbb{R}^m$, where $\mu : (\mathbb{C} - 0)^m \rightarrow \mathbb{R}^m$ is given by the formula

$$\mu(z_1, \dots, z_m) = (\log |z_1|, \dots, \log |z_m|)$$

and $A \subset (\mathbb{C} - 0)^m$ is the zero locus of p in the complement of the coordinate hyperplanes.

Let Δ be the Newton polytope of p , $T \supset (\mathbb{C} - 0)^m$ be the toric variety associated to Δ and $\mu_T : T \rightarrow \Delta$ be the moment map of T ; see e.g. [4]. Note that $\mu : (\mathbb{C} - 0)^m \rightarrow \mathbb{R}^m$ is just a reparametrization of $\mu_T|_{(\mathbb{C}-0)^m} : (\mathbb{C} - 0)^m \rightarrow \text{Int}\Delta \subset \mathbb{R}^m$, where the interior of Δ gets mapped to the whole \mathbb{R}^m . Let \bar{A} be the closure in T of $A \subset (\mathbb{C} - 0)^m \subset T$. The image $\mu_T(\bar{A}) \subset \Delta$ is called the compactified amoeba [4].

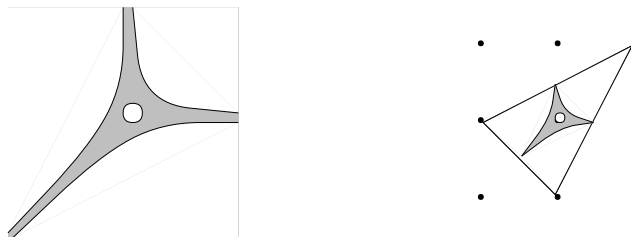


Figure 7. Amoeba $\mu(A)$ and compactified amoeba $\mu_T(\bar{A})$.

Consider the region $\mathbb{R}^m - \mu(A)$. It consists of bounded and unbounded components. It was observed in [4] that each component of $\mathbb{R}^m - \mu(A)$ is convex and that unbounded components of $\mathbb{R}^m - \mu(A)$ correspond (inductively) to the complementary regions of the amoebas of the intersection of \bar{A} with the toric subvarieties corresponding to the faces of Δ .

Remark 9. For $m = 2$ these intersections are zero-dimensional, so that the description of the unbounded components is easy. Indeed, the *tentacles* (i.e. the ends of $\mu(A)$) correspond to the intersection of \bar{A} with the axes l_j .

The number of such points counted with multiplicities is d_j , since d_j is the degree of the restriction of p to l_j . Thus, if \bar{A} intersects l_j transversely and no two intersection points are on the same fiber of $\bar{\mu}$, the unbounded components of $\mathbb{R}^2 - \mu(A)$ are in one-to-one correspondence with the integer points on the boundary of Δ , i.e. with the set $\partial\Delta \cap \mathbb{Z}^m$.

The number of components of $\mathbb{R}^m - \mu(A)$ depends on the coefficients of p and not just on Δ . However, Forsberg et al. [3] obtained the following upper bound for this number in terms of Δ .

THEOREM 4 ([3]). *There is a natural injective map ind from the set of components of $\mathbb{R}^m - \mu(A)$ to $\Delta \cap \mathbb{Z}^m$.*

There do exist amoebas with a bijective map ind ; see the appendix for the construction. In the proof of our main theorems we use a topological interpretation of ind .

3. Amoebas from a topological point of view

In this section we give a topological proof of Theorem 4. Also we define the *logarithmic Gauss map* γ for \bar{A} and compute its degree.

3.1. Proof of Theorem 4. If $x \in \mathbb{R}^m - \mu(A)$ then $\mu^{-1}(x)$ is an m -dimensional torus in $(\mathbb{C} - 0)^m$ not intersecting the hypersurface A . Therefore the linking number with the closure of A in \mathbb{C}^m produces a well-defined linear function $\text{lk} : H_1(\mu^{-1}(A)) \rightarrow \mathbb{Z}$. But $H_1(\mu^{-1}(A)) = \mathbb{Z}^m$, where the identification is given by coordinates in \mathbb{C}^m . Therefore, lk is given by m integer numbers (i_1, \dots, i_m) . Define $\text{ind} : \mathbb{R}^m - \mu(A) \rightarrow \Delta \cap \mathbb{Z}^m$ by $\text{ind}(x) = (i_1, \dots, i_m)$. Clearly, ind is locally constant and therefore defines a map on the set components of $\mathbb{R}^m - \mu(A)$. We need to prove that this map is injective and lands on $\Delta \cap \mathbb{Z}^m$.

Denote by $\pi : \mathbb{Z}^m \rightarrow \mathbb{Z}$ the projection defined by

$$\pi(j_1, \dots, j_m) = k_1 j_1 + \dots + k_m j_m.$$

To prove that $\text{ind}(x) \in \Delta$ it suffices to prove that $\pi(\text{ind}(x)) \in \pi(\Delta)$ for any $(k_1, \dots, k_m) \in \mathbb{Z}^m$.

Let $C \subset \mathbb{C}^m$ be the curve given by the parametrization $z_j = c_j t^{k_j}$, $0 \neq c_j \in \mathbb{C}$. For a generic choice of c_j , the pull-back of p to C is a polynomial in one variable whose Newton polytope is $\pi(\Delta)$. But $\mu|_{C - (0, \dots, 0)}$ is a circle fibration over the line $x_j = k_j t + \log |c_j|$. For $z \in C$ the circle $C \cap \mu^{-1}(\mu(z))$ represents the homology class $(k_1, \dots, k_m) \in H_1(\mu^{-1}(\mu(z)))$. The linking number in \mathbb{C}^m of $C \cap \mu^{-1}(\mu(z))$ and A is d , if $z \in C$ is sufficiently close to $(0, \dots, 0)$, D , if $z \in C$ is sufficiently close to infinity and anything in between for other choices of z , where $[d, D] = \pi(\Delta) \subset \mathbb{Z}$. This holds since we may use a part of C as a

membrane to compute the linking number. This implies that $\text{ind}(x) \in \Delta$ since the line $\mu(C)$ passes through the component of $\mathbb{R}^m - \mu(A)$ containing x for a suitable choice of c_j .

To show the injectivity of ind choose a line $l \subset \mathbb{R}^n$ with a rational slope (i.e. given by $x_j = k_j + b_j$ for some $k_j \in \mathbb{Z}$ and $b_j \in \mathbb{R}$) which passes through any pair of components in $\mathbb{R}^m - \mu(A)$. Let $x, y \in l$ be two points in different components of $\mathbb{R}^m - \mu(A)$. Since $\mu^{-1}[x, y] \cap A \neq \emptyset$ we may choose $c_j \in \mathbb{C}$ so that $\mu(C) = l$ and $C \cap \mu^{-1}[x, y] \cap A \neq \emptyset$ for C parametrized by $z_j = c_j t^{k_j}$. Therefore $\pi(\text{ind}(x)) \neq \pi(\text{ind}(y))$ and π is injective.

Note that the same argument also proves the convexity of components of $\mathbb{R}^m - \mu(A)$. Convexity holds even locally as the following lemma shows.

LEMMA 1. *Let $z \in A$ be a critical point of $\mu|_A$, $U \ni f(z)$ be a convex neighborhood of $f(z)$ in \mathbb{R}^2 and V be the component of $(\mu_A)^{-1}(U)$ which contains z . Then each component of $U - \mu(V)$ is convex.*

Proof. If not then there exists a straight closed interval $I \subset U \subset \mathbb{R}^2$ with both endpoints in the same component of $U - \mu(V)$ which intersects V . We may assume that I has a rational slope and that I is transverse to $\mu|_A$. Since U is contractible, $V \cap \mu^{-1}(I)$ is null-homologous in V and, therefore, null-homologous in $(\mathbb{C} - 0)^m$. But since the slope of I is rational there exists a holomorphic annulus Z which projects properly to I and intersects V . This leads to a contradiction. The intersection number of Z and $V \cap \mu^{-1}(I)$ in $\mu^{-1}(I)$ is positive on one hand, since V and Z are holomorphic, and zero on the other hand, since $V \cap \mu^{-1}(I)$ is null-homologous in $(\mathbb{C} - 0)^m$ and, therefore, null-homologous in $\mu^{-1}(I)$. □

3.2. *The logarithmic Gauss map* (cf. [9]). Suppose that $\bar{A} \subset T$ is nonsingular. Let $\gamma : A \rightarrow \mathbb{C}P^{m-1}$ be the map defined by

$$\gamma(z_1, \dots, z_m) = [z_1 \frac{\partial p}{\partial z_1}(z_1, \dots, z_m) : \dots : z_m \frac{\partial p}{\partial z_m}(z_1, \dots, z_m)].$$

The geometric description of γ is the following. Let $z \in A$ and $U \ni z$ be a small neighborhood of z in $(\mathbb{C} - 0)^m$. Choose a branch of the holomorphic logarithm $\log_U : U \rightarrow \mathbb{C}^m, (z_1, \dots, z_m)$ and apply the Gauss map G to the image of A (i.e. map each point $\log_U(A \cap U)$ to the tangent hyperplane at that point). The composition $G \circ \log_U$ does not depend on the choice of the branch of the logarithm and gives γ .

Suppose that for any face of Δ the toric subvariety corresponding to that face intersects \bar{A} transversely. In this case γ extends to \bar{A} . Let us extend γ to $\bar{A} \cap \mu^{-1}(\text{int} \Delta')$ for any facet $\Delta' \subset \Delta$. Without loss of generality we may assume that the hyperplane corresponding to Δ' is $x_m = 0$; otherwise we change the

coordinates in $(\mathbb{C} - 0)^m$ by $Z_j = z_1^{b_{j1}} \dots z_m^{b_{jm}}$ for a suitable integer b_{jk} . Denote by $p_{\Delta'} = \sum_{\Delta'} a_{j_1, \dots, j_m} z_1^{j_1} \dots z_m^{j_m}$ the truncation of $p = \sum_{\Delta} a_{j_1, \dots, j_m} z_1^{j_1} \dots z_m^{j_m}$ to Δ' . Then $p_{\Delta'}$ is a polynomial in $(m - 1)$ variables z_1, \dots, z_{m-1} . Define

$$\bar{\gamma}(z_1, \dots, z_{m-1}, 0) = [z_1 \frac{\partial p_{\Delta'}}{\partial z_1}(z_1, \dots, z_{m-1}) : \dots : z_{m-1} \frac{\partial p_{\Delta'}}{\partial z_{m-1}}(z_1, \dots, z_{m-1}) : 0]$$

for $(z_1, \dots, z_{m-1}, 0) \in \bar{A} \cap \mu^{-1}(\text{int}\Delta')$. Inductively by codimension of the faces of Δ , γ extends to a holomorphic map

$$\bar{\gamma} : \bar{A} \rightarrow \mathbb{C}P^{m-1}$$

between closed manifolds of the same dimension.

LEMMA 2. *The degree of $\bar{\gamma}$ is $n! \text{Vol}\Delta$.*

Proof. The inverse image $\bar{\gamma}^{-1}([0 : \dots : 0 : 1])$ is the zero set of $z_j \frac{\partial p}{\partial z_j}$, $j = 1, \dots, m - 1$, and p . The Newton polytope of each of these polynomials is Δ . By Kouchnirenko’s theorem [11] the number of points in $\bar{\gamma}^{-1}([0 : \dots : 0 : 1])$ counted with multiplicities is $n! \text{Vol}\Delta$. □

4. Real two-dimensional amoebas

Suppose now that the coefficients of p are real and $m = 2$. Then $\mu|_A : A \rightarrow \mathbb{R}^2$ is a map between smooth surfaces. Its generic singularities in the class of smooth maps are folds and cusps. Denote by $F \subset A$ the locus of critical points of $\mu|_A$, i.e. the points where μ_A is not submersive.

LEMMA 3. $F = \gamma^{-1}(\mathbb{R}P^1)$.

Proof. The holomorphic (multivalued) logarithm maps the fiber tori $\mu^{-1}(x)$, $x \in \mathbb{R}^2$ to the purely imaginary planes $\{\text{Re}z_1 = y_1, \text{Re}z_2 = y_2\}$ in \mathbb{C}^2 . Therefore, $z \in A$ is a critical point of μ_A if and only if the logarithmic image of the tangent plane at z contains a purely imaginary vector. But this holds if and only if $\gamma(z) \in \mathbb{R}P^1 \subset \mathbb{C}P^1$. □

Denote by $\mathbb{R}A \subset A$ the set of real zeroes of p in \mathbb{R}^2 and by $\mathbb{R}\bar{A} \subset \bar{A}$ its compactification in $\mathbb{R}T$.

COROLLARY 4. $\mathbb{R}A \subset F$.

Indeed, $\gamma(\mathbb{R}A) \subset \mathbb{R}P^1$ from the definition of γ . However, $\mathbb{R}A$ does not always coincide with F ; e.g. $\mathbb{R}A$ may be empty while F , being the folds of a proper degree 0 map $A \rightarrow \mathbb{R}^2$, is never empty. If a point z belongs to F then

its conjugate point \bar{z} also belongs to F . Thus, imaginary folds are double folds and the number of points in the inverse image of μ jumps by four at those folds.



Figure 8. Real folding and imaginary double folding.

Remark 10. Besides folds and cusps, which are stable singularities for smooth maps between surfaces there are two new stable singularities for μ pictured in Figure 9. These singularities persist under small real perturbations of p but decompose into folds and cusps under small imaginary perturbations. Indeed, $\bar{\gamma} : \bar{A} \rightarrow \mathbb{C}P^1$ is a branched covering. Branching points are the points of logarithmic inflection, i.e. inflection after taking the holomorphic logarithm. Generically, the branching is of multiplicity 2. The new singularities correspond to the case when the branching points are real. By Lemma 3, F has double points at the double branching points of γ . Only one of the branches at a double point of F may be from $\mathbb{R}A$. If the image of the other one under μ is not constant then it corresponds to the imaginary double folds pictured in Figure 9 on the right. If it is constant then a whole circle in F is mapped to a point and this circle must have another real point where the circle meet another branch of $\mathbb{R}A$. This is a pinching singularity pictured in Figure 9 on the left. It is a double point of $\mu(\mathbb{R}A)$ and both branches have an inflection at this point. The next two examples show that both cases appear.



Figure 9. A pinching and a junction of real and imaginary folds.

Example 1. Let $p(x, y) = xy - x - y + a$, where $0 < a < 1$. The corresponding hyperbola and its amoeba are pictured in Figure 10.

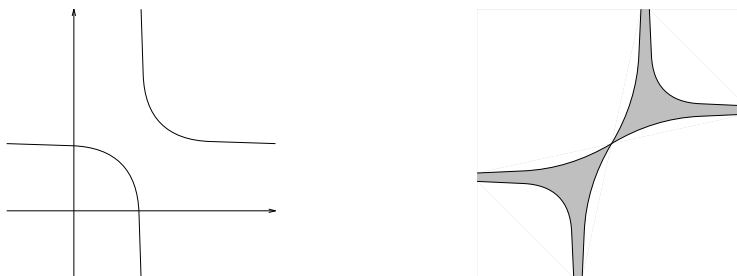


Figure 10. A hyperbola and its amoeba.

Indeed, the curve A is defined by $y - 1 = \frac{1-a}{x-1}$. The image of the circle $|x| = c_x$ under $\frac{1-a}{x-1}$ is a circle which intersects the circle $|y| = c_y$ in two points if $(c_x, c_y) \in \mu(A) - \mu(\mathbb{R}A)$, is tangent to it if $(c_x, c_y) \in \mu(\mathbb{R}A)$, is disjoint to it if $(c_x, c_y) \notin \mu(A)$ and coincides with it if $c_x = c_y = \sqrt{a}$.

Example 2. Let $p(x, y) = y - x^2 + 2x - a$, $a > 1$. The corresponding parabola and its amoeba are pictured in Figure 11. Indeed, $\mathbb{R}T$ in this case is a weighted projective space with weights $(1, 2, 1)$ and $\mathbb{R}A$ intersects the x -axis in a pair of conjugate imaginary points. The double imaginary folding must merge to the real folding (by Lemma 2 the degree of γ is 2).

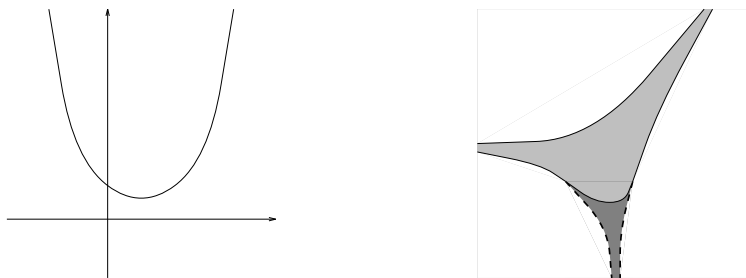


Figure 11. A parabola and its amoeba.

5. Proof of the main theorems

5.1. *Proof of Theorems 1 and 3.* Theorem 1 is a special case of Theorem 3, where $T = \mathbb{C}P^2$. All the lemmas in this subsection are formulated under the hypothesis of Theorem 3.

LEMMA 5. $F = \mathbb{R}A$.

Proof. By Pick’s formula, twice the area of Δ is equal to twice the number of lattice points g inside Δ plus the number of lattice points h on the boundary minus 2. Recall that g is the genus of $\mathbb{C}A$ (see [10]). By the maximality assumption (Definition 2) g is the number of closed components (ovals) of $\mathbb{R}A$. Note that for any oval C the inverse image $(\gamma|_C)^{-1}(x)$, $x \in \mathbb{R}P^1$, consists at least of two points. Indeed, $\mu(C) \subset \mathbb{R}^2$ is also an oval and its projection to \mathbb{R} along the direction determined by x must have two endpoints.

By definition of d_j we have $h = \sum_j d_j$. By the maximality assumption (Definition 2) the number of nonclosed components (arcs) of $\mathbb{R}A$ both of whose endpoints belong to the same axis l_j of T is $\sum_j (d_j - 1) = h - n$. Choose $x \in \mathbb{R}P^1$ to be any point which does not correspond to the orthogonal direction of the slope of a side of Δ . Then for any such arc B the inverse image $(\gamma|_B)^{-1}(x)$, $x \in \mathbb{R}P^1$, consists at least of one point.

Denote by D the n remaining arcs. The cyclical maximality assumption implies that $(\gamma|_D)^{-1}(x)$ consists at least of $n - 2$ points (the sum of the angles of a Euclidean n -gon is $\pi(n - 2)$).

Adding the above we conclude that $(\gamma|_D)^{-1}(x)$ consists at least of $2\text{Area}\Delta$ points. By Lemma 2 it consists precisely of $2\text{Area}\Delta$ points and $\gamma^{-1}(\mathbb{R}P^1) = F = \mathbb{R}A$. \square

COROLLARY 6. $\mu(\mathbb{R}A)$ does not have inflection points.

Proof. An inflection point of $\mu(\mathbb{R}A)$ would correspond to a real critical point of γ . By Lemma 3 that would correspond to a singular point of F . But $\mathbb{R}A$ is nonsingular and thus Lemma 5 yields a contradiction. \square

COROLLARY 7. The order of the intersection points $l_j \cap c_j$ on l_j and on c_j (see Definition 2) agrees.

Proof. If these orders do not coincide then one of the arcs of $\mu(\mathbb{R}A)$ must have an inflection point since the ends of $\mu(\mathbb{R}A)$ are convex in the half-planes cut by the asymptotes. \square

LEMMA 8. $\partial\mu(A) = \mu(\mathbb{R}A)$.

Proof. Lemma 5 implies that $\partial\mu(A) \subset \mu(\mathbb{R}A)$. Suppose that $\partial\mu(A) \neq \mu(\mathbb{R}A)$. Then for some $x \in \mathbb{R}^2$, $x \notin F$, $\mu^{-1}(x)$ consists of more than 2 points. Let L be a line passing through x with a rational slope which is not orthogonal to the slope of a side of Δ . Let $y \in L$ be a point close to infinity in L so that $\mu^{-1}(y) = \emptyset$. By Corollary 6 each component of $\mathbb{R}A$ cuts \mathbb{R}^2 into a convex and a nonconvex half. We call the convex half *the interior* of the component (even if this component is noncompact). By Lemma 1, if x belongs to the interior of a components and y belongs to the interior of b components then the number of points in $\mu^{-1}(x)$ is $2(b - a)$. But $b = 1$ and because of the maximality there is only one arc which joins the sides of Δ adjacent to y and only the interior of this arc may contain x . Therefore, $2(b - a) \leq 2$. \square

COROLLARY 9. $\mu|_{\mathbb{R}A}$ is an embedding.

Proof. A double point of $\mu(\mathbb{R}A)$ cannot be an inflection point of a branch by Corollary 6. Therefore, the other branch must intersect $\text{Int}\mu(A)$. This gives a contradiction to Lemma 8. \square

COROLLARY 10. *For any $(i_1, i_2) \in \Delta$ there exists a unique component Ω_{i_1, i_2} of $\mathbb{R}^2 - \mu(A)$. This component is a disk bounded by an oval of $\mu(\mathbb{R}A)$ if $(i_1, i_2) \in \text{Int}\Delta$ and a half-plane bounded by an arc of $\mu(\mathbb{R}A)$ if $(i_1, i_2) \in \partial\Delta$. Any component of $\mathbb{R}^2 - \mu(A)$ is Ω_{i_1, i_2} , $(i_1, i_2) \in \Delta$.*

In other words, the amoeba $\mu(A)$ is isotopic to the amoeba pictured in Figure 12.

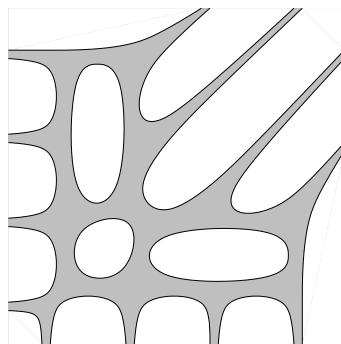


Figure 12. Amoeba of a curve in cyclically maximal position in $\mathbb{R}T$.

Proof. The component Ω_{i_1, i_2} is unique by Theorem 4. By the same theorem, the value of ind on any component of $\mathbb{R}^2 - \mu(A)$ belongs to Δ . But by Lemma 8 any component of $\mu(\mathbb{R}A)$ belongs to the boundary of a component in $\mathbb{R}^2 - \mu(A)$; because of maximality of $\mathbb{R}\bar{A}$ the number of components of $\mathbb{R}A$ is equal to the number of lattice points in Δ . Thus, we have a disk $\Omega(i_1, i_2)$ for any $(i_1, i_2) \in \text{Int}\Delta$ and a half-plane for any $(i_1, i_2) \in \partial\Delta$. \square

To reconstruct the topological type of $(\mathbb{R}T, \mathbb{R}\bar{A})$ it now suffices to know the distribution of the components of $\mathbb{R}A$ among the quadrants of $(\mathbb{R} - 0)^2$. The following lemma determines the lifting of components of $\mu(\mathbb{R}A)$ to $(\mathbb{R} - 0)^2 = \mu^{-1}(\mathbb{R}^2) \cap \mathbb{R}T$ and finishes the proof of Theorem 3 and Theorem 1.

Denote $\mathbb{R}_{(-1)^{j_1}, (-1)^{j_2}}^2 = \{(x_1, x_2) \mid (-1)^{j_1}x_1 > 0, (-1)^{j_2}x_2 > 0\}$. Fix a point $y \in \mathbb{R}A$; then $\mu(y) \in \partial\Omega_{j_1, j_2}$ for some $(j_1, j_2) \in \Delta$. Without loss of generality we may assume that $y \in \mathbb{R}_{(-1)^{j_2}, (-1)^{j_1}}^2$ (otherwise we change the signs of some of the coordinates). It turns out that, after this choice of signs, the same is true for all $\mathbb{R}A \cap \partial\mu(A)$.

LEMMA 11. *If $x \in \mathbb{R}A$ and $\mu(x) \in \partial\Omega_{i_1, i_2}$ then $x \in \mathbb{R}_{(-1)^{i_2}, (-1)^{i_1}}^2$.*

Proof. Connect $\mu(x)$ and $\mu(y)$ with a smooth path Q inside $\mu(A)$. The homology class of the circle $C = (\mu|_A)^{-1}(Q)$ in $H_1((\mathbb{C} - 0)^2) = \mathbb{Z}^2$ is $(j_2 - i_2, j_1 - i_1)$. To see this consider two loops $\alpha \subset \mu^{-1}(\mu(x))$ and $\beta \subset \mu^{-1}(\mu(y))$ not intersecting A which represent the same homology class $(a, b) \in H_1((\mathbb{C} - 0)^2)$.

The difference of the linking number of α and β with the closure of A in \mathbb{C}^2 is equal to $(j_1 - i_1)a - (j_2 - i_2)b$. But on the other hand it must be equal to the intersection number of an annulus, connecting α and β , with A . And that number is equal, in turn, to the intersection number in $\mu^{-1}(\mu(x))$ of α and the projection of C (note that the real 2-torus $\mu^{-1}(\mu(x))$ is a deformational retract of $(\mathbb{C} - 0)^2$).

Let R be a path connecting x and y in $\mathbb{R}^2 \supset (\mathbb{R} - 0)^2$. Points x and y separate C into two half-circles. Let D be a membrane in \mathbb{C}^2 spanned by the union of R and a half-circle of C . Then $D \cup \text{conj}(D)$ is a membrane for C . The parity of the intersection number of D with the x_k -axis in \mathbb{C}^2 ($k = 1, 2$) coincides with the parity of the intersection number of R with the real part of the x_k -axis since all imaginary intersection points come in pairs. Thus, the intersection number is odd if and only if x and y are separated by the x_k axis in \mathbb{R}^2 . But $[C] = (j_2 - i_2, j_1 - i_1) \in H_1((\mathbb{C} - 0)^2)$ and thus the linking number of C and the x_k -axis is $j_{3-k} - i_{3-k}$. \square

5.2. *Proof of Theorem 2.* By Remark 3 we may assume that $d \geq 4$. We deduce Theorem 2 from Theorem 1. Indeed, if $\mathbb{R}\bar{A}$ is in maximal position with respect to l_1, l_2, l_3, l_4 then it is in maximal position with respect to l_1, l_2, l_3 and thus of the type of Theorem 1. Suppose without loss of generality that the d intersection points with l_4 belong to the arc C of $\mathbb{R}\bar{A} - (l_1 \cup l_2 \cup l_3)$ connecting l_1 and l_2 . Since $\mathbb{R}\bar{A}$ is in maximal position with respect to l_1, l_2, l_4 and $d \geq 4$ the region between C and the corner between l_1 and l_2 must contain an oval of $\mathbb{R}\bar{A}$ by Theorem 1. But that is impossible by Theorem 1 applied to $\mathbb{R}\bar{A}$ and l_1, l_2, l_3 .

Appendix. Viro’s patchworking and amoebas

Viro introduced a patchworking technique for constructing algebraic curves in his dissertation. This construction is described in a very elementary way in [8]. It turns out that the same construction also provides patchworking for amoebas.

Let us recall the construction. Let Δ be a convex lattice polygon in \mathbb{R}^2 , T be the toric surface corresponding to Δ and $\mu_T : T \rightarrow \Delta$ be the moment map. Any function $h : \Delta \cap \mathbb{Z}^2 \rightarrow \mathbb{Z}$ determines a subdivision of Δ into a union of smaller polygons Δ_j by the following rule. Let R be the convex hull of $\{(x, y, t) \in \mathbb{Z}^3 \mid (x, y) \in \Delta, t \geq h(x, y)\}$. The region R is a semi-infinite polyhedral region. We define Δ_j to be the vertical projections of its faces.

Let $\{p_j = \sum_{\Delta_j} a_{m,n} x^m y^n\}$ be a collection of real polynomials such that the Newton polygon of p_j is Δ_j . We assume that the truncations of p_j and p_k to any common face of Δ_j and Δ_k coincide. Let Γ be a side of Δ_j for some j .

The truncation p_Γ is a weighted homogeneous polynomial in two variables. Suppose that it does not have multiple roots for any Γ and all its roots are real.

Let T_j be the toric surface associated to Δ_j and let $\mu_j : T_j \rightarrow \Delta_j$ be the moment map. Suppose that p_j defines a nonsingular curve A_j in T_j . Suppose that the only singularities of $\mu_j|_{A_j}$ are real folds, imaginary double folds, double cusps and the two singularities from Remark 9. Note that any cusp point of $\mu_j|_{A_j}$ is imaginary and thus, together with the conjugate point, it necessarily forms a double cusp. Suppose that the images of the folds $\mu_j(F_j) \subset \text{Int}\Delta_j$ of μ_j do not have self-tangencies or triple points.

Define the *patchworking* polynomial p_t of $\{p_j\}$ by

$$\sum_{\Delta} a_{m,n} t^{h(m,n)} x^m y^n$$

for $t > 0$. The Newton polygon of p_t is Δ . Denote the zero set of p_t in $(\mathbb{C} - 0)^2$ with A and its zero set in the toric surface T corresponding to Δ by \bar{A} . Let F be the folds of the moment map $\mu|_A$.

The curve $\bigcup_j \mu_j(F_j)$ is not properly embedded in $\text{Int}\Delta$. To make it proper we need to connect the loose ends on both sides of Γ_k for each interior edge Γ_k . We have two pairs of those ends for each zero of the truncation $p|_{\Gamma_k}$, one pair from each side of Γ_k . The two ends of each pair correspond to a pair of different quadrants of $(\mathbb{R} - 0)^2$. The two ends from the other pair correspond to the same pair of quadrants; thus we have a natural identification between the two pairs of ends. We connect the corresponding ends across Γ_k without introducing a self-intersection or introducing a new ordinary double point (depending on the mutual position of the two pairs). Let f be the resulting proper curve in $\text{Int}\Delta$. Its isotopy type is determined by the isotopy types of $\mu_j(F_j)$ and the distribution of their real ends among the components of $(\mathbb{R} - 0)^2$.

PROPOSITION A1. *The pairs $(\text{Int}\Delta, \mu_T(F))$ and $(\text{Int}\Delta, f)$ are homeomorphic. Under this homeomorphism the singularities of $\mu_j|_{A_j}$ correspond to the singularities of the same type of $\mu_T|_A$ and the new double points of f correspond to the pinching points of $\mu_T(A)$.*

Proof. Since R is convex there exists a coordinate change of the type $X = xt^a, Y = yt^b, a, b \in \mathbb{Z}$ for every Δ_j such that the pull-back P of p under this coordinate change is given by

$$\begin{aligned} (1) \quad P(X, Y) &= \sum_{\Delta} a_{m,n} t^{H(m,n)} X^m Y^n \\ &= t^H \left(\sum_{\Delta_j} a_{m,n} X^m Y^n + \sum_{\Delta - \Delta_j} a_{m,n} t^{H(m,n) - H} X^m Y^n \right), \end{aligned}$$

$H(m, n) \geq H$ and $H(m, n) = H$ if and only if $(m, n) \in \Delta_j$. On the other hand, the pull-back of p_j is

$$(2) \quad P_j(X, Y) = \sum_{\Delta_j} a_{m,n} t^{H(m,n)} X^m Y^n.$$

Let D_j be a bounded disk region in the (X, Y) -plane such that the amoeba of P_j is essentially contained in it; i.e., $\mu_j^{-1}(\mathbb{R}^2 - D_j) \cap \{(X, Y) \mid P(X, Y) = 0\}$ is a disjoint union of annuli which project under μ to strips connecting infinity with the D where the folds are the only singularities of the projection. Let K_j be the region corresponding to D_j in the (x, y) -plane. Note that K_j depends on the value of t .

If $t > 0$ is sufficiently small then the zero sets of p and p_j in $\mu^{-1}(K_j) \subset (\mathbb{C} - 0)^2$ are sufficiently close (cf. (1) and (2)). On the other hand for sufficiently small $t > 0$ and $j \neq l$, $K_j \cap K_l = \emptyset$. By assumption the singularities of $\mu|_{A_j}$ are from our list; they are stable under real deformation. Therefore, $\mu|_A$ has the same singularities over K_j as $\mu|_{A_j}$. By the other assumption $\mu(F_j)$ does not have double tangencies or triple points. Therefore, $\mu(F) \cap K_j$ and $\mu(F_j) \cap K_j$ are isotopic in K_j .

Note that we chose K_j so that $A_j - \mu^{-1}(K_j)$ is a collection of disjoint annuli. Recall that the genus of A is equal to the number of the lattice points in $\text{Int}\Delta$ and the same is true for A_j . This implies that $A - \bigcup_j \mu^{-1}(K_j)$ is a collection of disjoint annuli. By Lemma 3 the image of each of these annuli double covers a small disk around the orthogonal direction to Γ_j in $\mathbb{C}P^1$ under the logarithmic Gauss map. Therefore, this covering has two branching points. If they are real we have a pinching. If they are imaginary then the image of the folds of the annuli under μ is embedded. □

The curve obtained as a result of patchworking of polynomials whose Newton polygons Δ_j are triangles of area $\frac{1}{2}$ in \mathbb{R}^2 are called T-curves (see [8]). Note that any strictly convex function $h : \mathbb{R}^2 \rightarrow \mathbb{R}$ gives a decomposition of Δ into such triangles.

COROLLARY A2. *If A is a T-curve then for every $(m, n) \in \Delta \cap \mathbb{Z}^2$ there exists a component Ω of $\mathbb{R}^2 - \mu(A)$ such that $\text{ind}\Omega = (m, n)$.*

Indeed, Proposition A1 implies this even in a slightly more general situation when $\text{Ind}\Delta_j \cap \mathbb{Z}^2 = \emptyset$ for each j .

Let Δ_1 be the convex hull of (m_0, n_0) , (m_1, n_1) and (m_2, n_2) . Let Δ_2 be the convex hull of (m_1, n_1) , (m_2, n_2) and (m_3, n_3) . Suppose that $\text{Area}(\Delta_1) = \text{Area}(\Delta_2) = \frac{1}{2}$ and $\text{Int}\Delta_1 \cap \text{Int}\Delta_2 = \emptyset$. Suppose that Δ_1 and Δ_2 are contained in a patchworking of a Newton polygon Δ and $a_j \neq 0$ are the coefficients at $x^{m_j} y^{n_j}$. Let $\Gamma = \Delta_1 \cap \Delta_2$.

LEMMA A3. *If $a_1a_2 < 0$ then pinching at Γ occurs if and only if $a_0a_3 < 0$. If $a_1a_2 > 0$ and $(m_0, n_0) \equiv (m_2, n_2) \pmod{2}$ then pinching at Γ occurs if and only if $a_0a_3 < 0$. If $a_1a_2 > 0$ and $(m_0, n_0) \not\equiv (m_2, n_2) \pmod{2}$ then pinching at Γ occurs if and only if $a_0a_3 > 0$.*

Note that this lemma agrees with Example 1.

Proof. The pull-back of $a_jx^{m_j}y^{n_j}$ under the sign change of the x -coordinate is $(-1)^{m_j}x^{m_j}y^{n_j}$. Such a change does not affect amoeba. Thus, changing signs of x or y if needed we may assume that $a_1a_2 < 0$. In this case the curve $a_0x^{m_0}y^{n_0} + a_1x^{m_1}y^{n_1} + a_2x^{m_2}y^{n_2} = 0$ (which is just a reparametrization of a line in $(\mathbb{R} - 0)^2$, since $\text{Area}(\Delta_1) = \frac{1}{2}$) intersects the positive quadrant in $(\mathbb{R} - 0)^2$. The image of this intersection under μ does not have inflection points. Its curvature is directed towards (m_1, n_1) if $a_0a_2 > 0$ and towards (m_2, n_2) otherwise. The same is true for $a_3x^{m_3}y^{n_3} + a_1x^{m_1}y^{n_1} + a_2x^{m_2}y^{n_2} = 0$ in Δ_2 . Thus pinching occurs if and only if $a_0a_3 < 0$. \square

COROLLARY A4. *For any convex lattice polygon Δ there exists a polynomial p whose Newton polygon is Δ and the corresponding real curve $\mathbb{R}\bar{A} \subset \mathbb{R}T$ is in cyclically maximal position.*

Proof. We construct $\mathbb{R}\bar{A}$ as a T-curve. Decompose Δ into a union of triangles Δ_j of area $\frac{1}{2}$ using a strictly convex function h . Take $a_{m,n} = (-1)^{(m-1)(n-1)}$ (cf. [8]) as the coefficients of p_j . By Lemma A3, pinching does not occur. Thus, by Proposition A1, the corresponding curve is in cyclically maximal position in $\mathbb{R}T$. \square

HARVARD UNIVERSITY, CAMBRIDGE, MA
E-mail address: mihalkin@math.harvard.edu

REFERENCES

[1] V. I. ARNOLD, The situation of ovals of real plane algebraic curves, the involutions of four-dimensional smooth manifolds, and the arithmetic of integral quadratic forms, *Funkcional. Anal. i Priložen* **5** (1971), 1–9.
 [2] L. BRUSOTTI, Curve generatrici e curve aggregate nella costruzione di curve piane d'ordine assegnato dotate del massimo numero di circuiti, *Rend. Circ. Mat. Palermo* **42** (1917), 138–144.
 [3] M. FORSBERG, M. PASSARE, and A. TSIKH, Laurent determinants and arrangements of hyperplane amoebas, preprint, 1998.
 [4] I. M. GELFAND, M. M. KAPRANOV, and A.V. ZELEVINSKY, *Discriminants, Resultants, and Multidimensional Determinants*, Birkhäuser Inc., Boston, MA, 1994.
 [5] A. HARNACK, Über Vieltheiligkeit der ebenen algebraischen Curven, *Math. Ann.* **10** (1876), 189–199.
 [6] D. HILBERT, Über die reelen Züge algebraischer Curven, *Math. Ann.* **38** (1891), 115–138.
 [7] ———, *Mathematische Probleme*, *Arch. Math. Phys.* **1** (1901), 213–237.

- [8] I. ITENBERG and O. VIRO, Patchworking algebraic curves disproves the Ragsdale conjecture, *Math. Intelligencer* **18** (1996), 19–28.
- [9] M. M. KAPRANOV, A characterization of A -discriminantal hypersurfaces in terms of the logarithmic Gauss map, *Math. Ann.* **290** (1991), 277–285.
- [10] A. G. KHOVANSKII, *Newton polyhedra and toric varieties*, *Funkcional. Anal. i Priložen* **11** (1977), 56–64.
- [11] A. G. KOUCHNIRENKO, Polyèdres de Newton et nombres de Milnor, *Invent. Math.* **32** (1976), 1–31.
- [12] I. G. PETROVSKY, On the topology of real plane algebraic curves, *Ann. of Math.* **39** (1938), 187–209.
- [13] G. M. POLOTOVSKII, A catalogue of M -splitting curves of order six, *Dokl. Akad. Nauk SSSR* **236** (1977), 548–551.
- [14] V. A. ROHLIN, Congruences modulo 16 in Hilbert’s sixteenth problem, *Funkcional. Anal. i Priložen* **6** (1972), 58–64.
- [15] O. YA. VIRO, Real plane algebraic curves: constructions with controlled topology, *Leningrad Math. J.* **1** (1990), 1059–1134.

(Received December 17, 1998)