ANNALES
POLONICI MATHEMATICI
LXXV.2 (2000)

Regular analytic transformations of $\mathbb{R}^2$

by Joseph Gubeladze (Tbilisi)

Abstract. Existence of *loops* for non-injective regular analytic transformations of the real plane is shown. As an application, a criterion for injectivity of a regular analytic transformation of $\mathbb{R}^2$ in terms of the Jacobian and the first and second order partial derivatives is obtained. This criterion is new even in the special case of polynomial transformations.

1. Introduction. We define the notion of a loop and use it to derive properties of *non-injective* regular analytic mappings $F : \mathbb{R}^2 \to \mathbb{R}^2$. We observe that a *loop system* exists for any such mapping (Proposition 4.6). This observation seems to be of independent interest. Basing on Proposition 4.6, we obtain a criterion for injectivity of a *regular real-analytic* 2-map (Theorem 6.1). This criterion, in the special case of polynomial mappings, translates into the following claim (Theorem 6.2): a non-degenerate polynomial map $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a global diffeomorphism of the real plane if there exist $\lambda, \kappa > 0$ such that $\deg(F) \le \frac{3}{2}\lambda + 3$ and $j(F)_z \ge \kappa \|z\|^\lambda$ for all z with $\|z\|$ sufficiently large.

In particular, any cubic transformation of $\mathbb{R}^2$ with Jacobian separated from 0 (i.e. $>$ const > 0) is a global diffeomorphism. Recently it has been shown by Gwoździewicz [Gw] that *any* cubic polynomial transformation of $\mathbb{R}^2$ with non-vanishing Jacobian is a global diffeomorphism.

One can check that the Jacobian of Pinchuk's counterexample [P] to the Strong Real Jacobian Conjecture (for two variables) approaches zero along a certain algebraic curve, extending to infinity. Thus our result says that on the other extreme, when the Jacobian grows rapidly with certain rate at infinity, such counterexamples do not exist.

Thanks are due to Madhav Nori for discussions of details and for his proof of Lemma 3.2, and to Arno van den Essen for sending a copy of [Gw].

2000 *Mathematics Subject Classification*: 14E09, 14P05, 32C05.
Key words and phrases: regular analytic maps, Jacobian, injectivity.
Supported in part by CRDF grant #GM1-115.

2. Preliminaries. "Analytic mapping" always means a real-analytic mapping. Let $F = (f, g) : \mathbb{R}^2 \to \mathbb{R}^2$ denote an analytic mapping. We fix an orientation on $\mathbb{R}^2$. $\|z\|$ will refer to the standard Euclidean norm of $z \in \mathbb{R}^2$.

For $z = (x, y) \in \mathbb{R}^2$ we put

$$J(F)_z = \begin{pmatrix} \frac{\partial f}{\partial x}(z) & \frac{\partial f}{\partial y}(z) \\ \frac{\partial g}{\partial x}(z) & \frac{\partial g}{\partial y}(z) \end{pmatrix}$$

and $j(F)_z = \det J(F)_z$. In this paper F is called *regular* (or *non-degenerate*) if $j(F)_z > 0$ for all $z \in \mathbb{R}^2$

O will denote the origin of $\mathbb{R}^2$, i.e. $O = (0, 0)$.

For $r > 0$ we denote by C_r the circle in $\mathbb{R}^2$ with radius r and centre at O, while D_r denotes the closed disc bounded by C_r. We put $S^1 = C_1$. For $z \in \mathbb{R}^2$ we denote by $C_{z,r}$ and $D_{z,r}$ the circle $z + C_r$ and the disc $z + D_r$ respectively. An *arc* of some circle always means a closed arc, strictly contained in this circle and not degenerating to a single point. An arc, unless specified otherwise, always means that of a circle with centre at the origin.

A *curve* in $\mathbb{R}^2$ is just a homeomorphic image of the unit interval $[0, 1]$. An *analytic curve* in $\mathbb{R}^2$ is the image of $[0, 1]$ under some injective analytic map $\theta = (\theta_1, \theta_2) : [0, 1] \to \mathbb{R}^2$. In the special case when the derivatives of θ_1 and θ_2 do not have a common zero in $[0, 1]$ the curve is called *regular*. A point $(\theta_1(t), \theta_2(t))$ of an analytic curve $P \subset \mathbb{R}^2$ is called *singular* if $t \in [0, 1]$ is a common zero of the derivatives of θ_1 and θ_2. For any analytic curve P the set of its singular points is finite. Further, two analytic curves having the same end points and infinite intersections coincide. A *piecewise analytic curve* is defined as a curve which is the union of finitely many successive analytic curves $P_1, \ldots, P_n$ so that $P_i \cap P_{i+1}$ is an end point of both P_i and P_{i+1}, $i \in [1, n-1]$.

A *path* refers to a continuous mapping from $[0, 1]$, and a *closed path* to one from S^1.

$\#(A)$ denotes the number of elements in A. $\mathbb{N} = \{1, 2, \ldots\}$. For a subset $A \subset \mathbb{R}^2$, homeomorphic to a convex subset $B \subset \mathbb{R}^2$, $\operatorname{int}(A)$ refers to the corresponding image of the relative interior of B. Further notations are explained in the text.

3. Analytic background. A map $f : A \to B$ is called *almost injective* if $\#(f^{-1}(b)) < \infty$ for all $b \in B$ and $\#(f^{-1}(b)) \le 1$ for all but finitely many $b \in B$.

LEMMA 3.1. *Any non-degenerate analytic map $F : S^1 \to \mathbb{R}^2$ admits a factorization $F = \phi \circ \gamma$ into two non-degenerate analytic mappings, where γ is an n-fold analytic covering of S^1 for some uniquely determined natural number n and $\phi : S^1 \to \mathbb{R}^2$ is an almost injective analytic map.*

P r o o f. Define an equivalence relation on S^1 as follows: for $x, y \in S^1$ we write $x \sim y$ if and only if any neighbourhoods of x and y contain $x' \neq x$ and $y' \neq y$ respectively, such that $F(x') = F(y')$. Then straightforward arguments show that the function $\#([-]_\sim) : S^1 \to \mathbb{N} \cup \{\infty\}$ is finite and constant with a value, say, n. It follows easily that n is the desired number. ∎

LEMMA 3.2. *Assume* $F : \mathbb{R}^2 \to \mathbb{R}^2$ *is a regular analytic mapping. Then* $F|_{C_r} : C_r \to \mathbb{R}^2$ *is almost injective for any* $r > 0$.

P r o o f. By Lemma 3.1, $F|_{C_r}$ admits a factorization $F|_{C_r} = \phi \circ \gamma$ for some analytic n-fold covering $\gamma : S^1 \to S^1$ and some analytic almost injective $\phi : C_1 \to \mathbb{R}^2$. It suffices to show that $n = 1$.

Consider the function $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$\Psi(r, \theta) = F(re^{i\theta}), \qquad r, \theta \in \mathbb{R}\ (\mathbb{R}^2 = \mathbb{C}).$$

Since $j(F)_z > 0$, $z \in \mathbb{R}^2$, we have the well defined function

$$\frac{\partial \Psi}{\partial \theta} : \mathbb{R}^2 \setminus \{O\} \to \mathbb{R}^2 \setminus \{O\}.$$

For any $t > 0$ we have the closed path

$$\varrho_t = \frac{\partial \Psi}{\partial \theta}\bigg|_{C_t} : C_t \to \mathbb{R}^2 \setminus \{O\}.$$

Let $W(\varrho_t, O)$ denote the winding number of ϱ_t around the origin O. Since for any $z \in C_r$ the vector $\varrho_r(z)$ has the same direction as the oriented tangent line to $F(C_r)$ at $F(z)$ we see that

$$n \leq W(\varrho_r, O). \tag{$*$}$$

For $t, t' > 0$ distinct the closed paths ϱ_t and $\varrho_{t'}$ are homotopic in $\mathbb{R}^2 \setminus \{O\}$. Using the fact that the winding numbers (around O) of closed homotopic paths encircling O are the same ([F, p. 157]) we get $W(\varrho_r, O) = W(\varrho_\varepsilon, O)$, where ε is any small positive number. By the non-degeneracy condition $F : \mathbb{R}^2 \to \mathbb{R}^2$ is an orientation preserving local homeomorphism. But then $W(\varrho_\varepsilon, O) = 1$ for $\varepsilon > 0$ small enough. By $(*)$ we are done. ∎

Let $F : \mathbb{R}^2 \to \mathbb{R}^2$ be a regular analytic mapping. Put

$$A = \{(p, q) \in \mathbb{R}^2 \times \mathbb{R}^2 \mid p \neq q,\ F(p) = F(q),\ \|p\| = \|q\|\}.$$

Let $\pi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}^2$ be the projection onto the first factor.

LEMMA 3.3. (I) $\#(C_r \cap \pi(A)) < \infty$ *for any* $r > 0$.

(II) *For any* $z \in \mathbb{R}^2$ *there exists a disc* $D_{z,r}$ *such that* $\pi(A) \cap D_{z,r}$ *is one of the following sets*:

(a) $\emptyset$,

(b) $\{z\}$,

(c) *the union of a finite system of analytic curves containing* z, *having no other points of intersection pairwise*, *and regular outside* $\{z\}$.

Proof. (I) Assume $\#(C_r \cap \pi(A)) = \infty$. Then $F(C_r)$ has an infinite number of self-intersections. This contradicts Lemma 3.2.

(II) $A \subset \mathbb{R}^2 \times \mathbb{R}^2$ is an analytic subset. Observe that $\dim(A) = 1$ because $\dim(A) = 2$ would contradict (I). Now the classical Puiseux Theorem says that irreducible components of A locally admit an analytic parametrization. We are done because such local parametrizations are inherited by $\pi(A) \subset \mathbb{R}^2$. ∎

4. Loop systems. In what follows $F = (f, g) : \mathbb{R}^2 \to \mathbb{R}^2$ will always denote a regular analytic mapping. We let π and A be as in Lemma 3.3 and put $B = \pi(A)$.

Definition 4.1. An arc α of a circle C_r $(r > 0)$ is called an F-*loop* if $F(\alpha)$ is homeomorphic to S^1 and $F : \alpha \to F(\alpha)$ only identifies the end points.

The *height* of α, denoted by $\mathrm{ht}(\alpha)$, is the radius of the circle that contains α (i.e. $\mathrm{ht}(\alpha) = r$).

The end points of an F-loop α will be denoted by p_α and q_α. Thus $p_\alpha, q_\alpha \in B$ for any F-loop α. Moreover, we assume that p_α precedes q_α with respect to the orientation of the plane of reference. The point $F(p_\alpha) = F(q_\alpha)$ will be called a *base point* of the homeomorphic circle $F(\alpha)$.

One more notation: for an F-loop α we denote by $F(\alpha)^+$ the homeomorphic disc in $\mathbb{R}^2$ that is bounded by $F(\alpha)$.

Definition 4.2. Assume $0 < a < b$, $a \in \mathbb{R}$, $b \in \mathbb{R} \cup \{\infty\}$. A system $\{\alpha_t\}_{[a,b[}$ is called a *continuous* F-*loop system* if the following conditions are satisfied:

(1) α_t is an F-loop for any $t \in [a, b[$,

(2) $p_{\alpha_t}, q_{\alpha_t} : [a, b[\to \mathbb{R}^2$ are continuous functions,

(3) $F(\alpha_{t_2})$ is contained in the interior of $F(\alpha_{t_1})^+$ whenever $a \le t_1 < t_2 < b$,

(4) $\mathrm{ht}(\alpha_t) = t$.

Lemma 4.3. *Let* a *and* b *be as in Definition* 4.2 *and* $\{\alpha_t\}_{[a,b[}$ *be a continuous* F-*loop system. If* $b < \infty$ *then* $\lim_{t\to b} p_{\alpha_t}$ *and* $\lim_{t\to b} q_{\alpha_t}$ *exist and are different.*

Proof. Let p be any limit point of $\{p_{\alpha_t}\}_{[a,b[}$ such that $\|p\| = b$. The structural description of $\pi(A)$ near p, as given in Lemma 3.3(II), shows that $p = \lim_{t\to b} p_{\alpha_t}$.

If the two limits coincide then the map $F : \mathbb{R}^2 \to \mathbb{R}^2$ is not a local homeomorphism at the corresponding limit point, a contradiction. ∎

LEMMA 4.4. *Let* $\{\alpha_t\}_{[a,b[}$ *be a continuous* F*-loop system for some* $0 < a < b < \infty$. *Then the curves* $\{p_{\alpha_t}\}_{[a,b]}$ *and* $\{q_{\alpha_t}\}_{[a,b]}$ *are piecewise analytic*, *where* $p_{\alpha_b} = \lim_{t\to b} p_{\alpha_t}$ *and* $q_{\alpha_b} = \lim_{t\to b} q_{\alpha_t}$.

Proof. Since $p_{\alpha_b} \neq q_{\alpha_b}$ (Lemma 4.3) we have $p_{\alpha_b}, q_{\alpha_b} \in B$. We also have $p_{\alpha_t}, q_{\alpha_t} \in B$ for $t \in [a, b[$. Now the claim follows directly from Lemma 3.3(II). ∎

DEFINITION 4.5. For $0 < a < \infty$ a system $\{\alpha_t\}_{[a,\infty[}$ is called an *F-loop system* if there is a sequence $a = a_0 < a_1 < a_2 < \ldots$ such that

(1) $\{\alpha_t\}_{[a_i,a_{i+1}[}$ is a continuous F-loop system for any $i = 0, 1, 2, \ldots$,

(2) $F(\alpha_{t_2})$ is contained in the interior of $F(\alpha_{t_1})^+$ whenever $a \le t_1 < t_2 < \infty$,

(3) either the sequence $a_0, a_1, \ldots$ is finite or $\lim_{i\to\infty} a_i = \infty$.

The main observation on loop systems is

PROPOSITION 4.6. *The map F is non-injective if and only if there exists an F-loop system.*

The proof of this claim is a lengthy sequence of mostly standard facts on the topology of $\mathbb{R}^2$ and local triviality of the analytic set $\pi(A)$ as in Lemma 3.3. We therefore skip the details and only sketch the course of proof: we start with a small circle C such that F maps it to a diffeomorphic circle. By blowing up C homothetically we reach the first position when $F(C)$ touches itself. Then C contains an F-loop $\alpha \subset C$. Moreover, the loop is *regular* in the sense that there is an intermediate arc $\alpha \subset \beta \subset C$ such that $F(\beta \setminus \alpha)$ does not intersect the homeomorphic disc bounded by $F(\alpha)$. Next we show that there is a continuous F-loop system $\{\alpha_t\}_{[a,b[}$ with $\alpha_a = \alpha$. By the Zorn lemma we can choose a maximal such system (w.r.t. the natural partial order). Assume $b < \infty$. The next crucial fact is the existence of a regular F-loop α' of height b which is mapped into the interiors of all the homeomorphic discs bounded by the $F(\alpha_t)$. We then iterate the process, and so on. The concluding step in the proof is showing that the resulting sequence $\{a, b, \ldots\}$ satisfies the condition 4.5(3).

5. Extension rate and curvature.

Assume $z \in \mathbb{R}^2 \setminus \{O\}$ and $r = \|z\|$. We define the *extension rate* of F at z as follows. Consider a small arc of C_r, say γ, that contains z in its interior. For $\varepsilon > 0$ we let γ_ε denote the polar projection of γ into $C_{r+\varepsilon}$. Then the normal to $F(\gamma)$ at $F(z)$ intersects $F(\gamma_\varepsilon)$ in a single point, providing ε is small enough. We denote this point by $F(z)_\varepsilon$. Now consider an infinitesimal square one of whose edges is tangent to C_r at z so that the square is outside the disc D_r. Its image under F is an infinitesimal parallelogram, tangent to $F(\gamma)$ at $F(z)$. This parallelogram is the image of the above mentioned tangent square under the derivative of F at z. The height of the parallelogram above the edge tangent to $F(\gamma)$ is $\varepsilon\|F(z)_\varepsilon - F(z)\| + o(\varepsilon)$. The limit

$$\text{e.r.}(F)_z = \lim_{\varepsilon\to 0} \frac{\|F(z)_\varepsilon - F(z)\|}{\varepsilon} > 0$$

will be called the *extension rate* of F at z.

We shall say that F has a *negative curvature* at z if for small $\varepsilon > 0$ the point $F(z)_\varepsilon$ and the centre of curvature of $F(\gamma)$ at $F(z)$ lie on the same ray with origin at $F(z)$. If these two points are separated by $F(z)$ then we say that F has a *positive curvature* at z. The *curvature* of F at z, denoted by $c(F)_z$, is defined as the real number whose absolute value equals the curvature of $F(\gamma)$ at $F(z)$ and whose sign is chosen as above.

THEOREM 5.1. *Let* $\{\alpha_t\}_{[a,\infty[}$ *be an* F*-loop system and let* $\{c_i\}_{i\in\mathbb{N}}$ *be any increasing sequence of positive numbers such that* $\lim_{i\to\infty} c_i = \infty$. *Then there exist* $\varrho > 0$, *a sequence* $\{t_i\}_{i\in\mathbb{N}} \subset [a,\infty[$ *and points* $z_i \in \alpha_{t_i}$ *such that*:

(A) $\text{e.r.}(F)_{z_i} < 1/(c_i t_i)$,
(B) $-c(F)_{z_i} > 1/\varrho$,
(C) $t_1 < t_2 < \ldots$ *and* $t_i \to \infty$ *as* $i \to \infty$.

We will need the following (easily checked) fact:

• For any compact subset $Z \subset \mathbb{R}^2$, $Z \neq \emptyset$, there exists a smallest disc containing Z. Moreover, the boundary of this smallest disc intersects Z at least in 2 points.

(Disc here means a set *congruent* to D_r for some $r \geq 0$.)

Proof of Theorem 5.1. First observe that it suffices to achieve (A)&(B).

Now for each $t \geq a$ we let D_{α_t} denote the smallest disc in $\mathbb{R}^2$ containing $F(\alpha_t)$. Let C_{α_t} denote the boundary of D_{α_t} and ϱ_{α_t} its radius. So the function $\varrho_{\alpha_t} : [a,\infty[\to\mathbb{R}_+$ is *strictly* decreasing. Put $\varrho = \varrho_{\alpha_a}$. We know that $\#(C_{\alpha_t} \cap F(\alpha_t)) \geq 2$. Clearly, either $\#(C_{\alpha_t} \cap F(\alpha_t)) < \infty$ or $C_{\alpha_t} = F(\alpha_t)$. The latter is excluded because we would have $C_{\alpha_t} = F(C_t)$. This is so because if an arc of C_t is mapped under F to the circle C_{α_t} then the image of the whole circle C_t cannot go outside C_{α_t} (we use the fact that F is regular). On the other hand the winding number of $F : C_t \to C_{\alpha_t}$ is at least 2 because already the proper arc $\alpha_t \subset C_t$ covers the whole circle C_{α_t} under the mapping F. In particular, $F|_{C_t}$ is not almost injective—a contradiction by Lemma 3.2.

It follows that

$$X_t = \alpha_t \cap F^{-1}((C_{\alpha_t} \cap F(\alpha_t)) \setminus \{F(p_{\alpha_t})\})$$

is a *finite* non-empty subset of $\mathrm{int}(\alpha_t)$. Now for each end point of α_t there are two possibilities:

(a) the tangent direction to C_t at this end point is mapped under F to a direction tangent to C_{α_t} at $F(p_{\alpha_t})$,

(b) the tangency as in (a) does not occur.

We put

$$\mathrm{Tan}_t = X_t \cup \{\text{the end points of } \alpha_t \text{ that satisfy the condition (a)}\}.$$

It is clear that $-c(F)_\xi = 1/\varrho_{\alpha_t} \geq 1/\varrho$ for each $\xi \in \mathrm{Tan}_t$.

Theorem 5.1 will clearly be proved once we show the following

CLAIM. *For any $\varepsilon > 0$ there exists $t \in [a, \infty[$ such that* $\min\{\text{e.r.}(F)_\xi \mid \xi \in \mathrm{Tan}_t\} < \varepsilon/t$.

First one convention: for a disc $D \subset \mathbb{R}^2$ and $\lambda > 0$ we denote by $\lambda \times D$ the image of D under the homothety centred at the centre of D with factor λ.

Assume to the contrary that there exists $\varepsilon > 0$ such that

$$\min\{\text{e.r.}(F)_\xi \mid \xi \in \mathrm{Tan}_t\} \geq \varepsilon/t$$

for any $t \in [a, \infty[$. Fix such an ε. For each t there are two possibilities: either (1) $F(p_{\alpha_t}) \notin C_{\alpha_t}$, or (2) $F(p_{\alpha_t}) \in C_{\alpha_t}$.

Let $0 < \kappa < 1$. Since the extension rate of F at a point of α_t represents a "shrinking rate" of the homeomorphic circle $F(\alpha_t)$ at the corresponding point, our lower bound for $\text{e.r.}(F)_\xi$ in case (1) implies the following: there exists $\delta_t > 0$ for which

$$F(\alpha_\tau) \subset (1 - \kappa\varepsilon(\tau - t)/t) \times D_{\alpha_t}$$

for all $\tau \in [t, t + \delta_t]$. Similar standard analytic arguments show that, in case (2), there exists $\delta_t > 0$ for which $F(\alpha_\tau)$ is contained in the convex hull of

$$((1 - \kappa\varepsilon(\tau - t)/t) \times D_{\alpha_t}) \cup \{F(p_{\alpha_t})\}$$

whenever $\tau \in [t, t + \delta_t]$. But the above-mentioned hull is obviously included in a disc of radius

$$\left(1 - \frac{\kappa\varepsilon(\tau - t)}{2t}\right)\varrho_{\alpha_t},$$

namely the one contained in D_{α_t} with boundary tangent to C_{α_t} at $F(p_{\alpha_t})$.

By integration we get

$$\varrho_{\alpha_t} \leq \left(1 - \frac{\kappa\varepsilon}{2}\int_a^t \frac{1}{\tau}\, d\tau\right)\varrho$$

for all $t \in [a, \infty[$. But the latter inequality implies $\varrho_{\alpha_t} < 0$ for t sufficiently large, which is absurd. ∎

Corollary 5.2. *If* $F:\mathbb{R}^2\to\mathbb{R}^2$ *is not injective then for any increasing sequence* $\{c_i\}_{i\in\mathbb{N}}$ *of positive numbers such that* $\lim_{i\to\infty} c_i=\infty$, *there exist* $\varrho,\sigma>0$ *and a system* $\{z_i\}_{i\in\mathbb{N}}\subset\mathbb{R}^2\setminus\{O\}$ *satisfying*:

(A) $\|z_1\|<\|z_2\|<\ldots$ *and* $\|z_i\|\to\infty$ *as* $i\to\infty$,
(B) $\mathrm{e.r.}(F)_{z_i}<1/(c_i\|z_i\|)$,
(C) $\|F(z_i)\|<\sigma$,
(D) $-c(F)_{z_i}>1/\varrho$.

Proof. This immediately follows from Proposition 4.6 and Theorem 5.1. ∎

Remark 5.3. The conditions (A)&(B) can actually be derived from Hadamard's classical results [Ha]. Moreover, the analyticity assumption on F can at this point be relaxed to a smooth local diffeomorphism. The new thing here is to have simultaneously the condition (D) satisfied.

6. Injectivity. For a regular analytic map $F:\mathbb{R}^2\to\mathbb{R}^2$ and $z\in\mathbb{R}^2$ we let $D^2(F)_z$ denote the maximum of the absolute values of all second order partial derivatives of F at z, and let $a(F)_z$ denote $\max_{v\in S^1}\|F'_z(v)\|$. Two applications of loops are as follows:

Theorem 6.1. *Let F be a regular analytic transformation of $\mathbb{R}^2$. Assume there exist $\kappa_1,\kappa_2,\lambda>0$ such that $\kappa_1\|z\|^\lambda\le j(F)_z$ and $D^2(F)_z a(F)_z\le\kappa_2\|z\|^{3\lambda+3}$ for $\|z\|$ large enough. Then F is injective.*

Theorem 6.2. *Let $F:\mathbb{R}^2\to\mathbb{R}^2$ be a non-degenerate polynomial mapping of degree d. Assume there exist $\kappa,\lambda>0$ such that $d\le\frac{3}{2}\lambda+3$ and $\kappa\|z\|^\lambda\le j(F)_z$ for $\|z\|$ large enough. Then F is a global diffeomorphism of $\mathbb{R}^2$.*

Remark. Any quadratic non-degenerate polynomial transformation of $\mathbb{R}^n$, $n\in\mathbb{N}$, is a global diffeomorphism [KR, §3].

Proof of Theorem 6.2. By [BR] it suffices to show that F is injective. We have $a(F)_z=\max_{v\in S^1}\|J(F)_z\cdot v^T\|$, where v^T is the column transpose. Thus $a(F)_z\le\|J(F)_z\|\cdot\|v^T\|=\|J(F)_z\|$, where $\|J(F)_z\|$ denotes the standard Euclidean norm of the matrix $J(F)_z$. In particular, there exists $\kappa'>0$ such that $a(F)_z\le\kappa'\|z\|^{d-1}$ for $\|z\|$ large enough. It is clear that $D^2(F)_z\le\kappa''\|z\|^{d-2}$ for some $\kappa''>0$ whenever $\|z\|$ is large enough. We get

$$d\le\tfrac{3}{2}\lambda+3 \;\Leftrightarrow\; (d-1)+(d-2)\le 3\lambda+3$$

and Theorem 6.1 applies. ∎

To prove Theorem 6.1 we need several inequalities.

First a few notations. We denote by vol the standard translation invariant volume function in $\mathbb{R}^2$. For two (measurable) subsets $M,N\subset\mathbb{R}^2$ we put

$$\mathrm{vol}(M,N)=\mathrm{vol}(M\setminus N)-\mathrm{vol}(M\cap N).$$

For $\varepsilon > 0$ and $M \subset \mathbb{R}^2$ we denote by εM the image of M under the homothety centred at O with factor ε.

For $a, b > 0$ an ellipse $E \subset \mathbb{R}^2$ will be called an (a,b)-*ellipse* if E is congruent to the ellipse $x^2/a^2 + y^2/b^2 = 1$ and is centred at O. For an (a,b)-ellipse E we denote by E_x the positive number such that $(E_x, 0) \in E$.

For an ellipse E the homeomorphic disc bounded by E will be denoted by E^+.

D^1 will refer to the disc $D_{(0,1),1}$.

Finally, for $z \in \mathbb{R}^2$ we denote by $D^{\mathrm{curv}}_{F,z}$ the disc centred at the curvature centre of $F(C_{\|z\|})$ at $F(z)$, with radius $|c(F)_z|$ ($D^{\mathrm{curv}}_{F,z}$ may be the whole halfplane).

We have the following three observations.

(1) *There exists a real number c_1 such that for any ellipse E centred at O, the inequality*

$$\frac{\mathrm{vol}(\varepsilon E^+, D^1)}{\varepsilon^3 E_x^3} > c_1$$

holds for $\varepsilon > 0$ small enough (*depending on E*).

In fact, by elementary geometric observations one concludes easily that $\mathrm{vol}(\varepsilon E^+, D^1)$ is *more* than the area of $\Sigma \setminus D^1$ modulo infinitesimals of higher order, where Σ is the triangle with vertices O, $(0,1)$, $(E_x, 0)$. Now the Taylor series expansion of arctan near 0 applies.

(2) *Let $0 < b \le d \le a$ and E be an (a,b)-ellipse, tangent to the line $y = d$. Then*

$$\frac{\pi}{4}\frac{ab}{d} < E_x.$$

It suffices to observe that E can be inscribed in the parallelogram having one edge on the line $y = d$ and containing $(E_x, 0)$ in its boundary. The area of this parallelogram is $4dE_x$ and the area of our ellipse is πab. Hence $\pi ab < 4dE_x$.

(3) *There is $c_2 > 0$ such that the F-image of any smooth path $P \subset D_{z,\varepsilon}$ is contained in the $c_2 D^2(F)_z \varepsilon^2$-neighbourhood of $F(z) + F'_z(P)$, providing ε is small enough* (*depending on z*). *Moreover, c_2 can be chosen so that the area of the above-mentioned neighbourhood is always at most $c_2 \Delta^2(F)_z a(F)_z l(P) \varepsilon^2$ for ε small enough* (*depending on z*), *where $l(P)$ is the length of P.*

(Here the image of a continuous mapping from $[0,1]$ (i.e. of a path) is itself called a path.)

Summing up (1), (2) and (3), and using the equalities

$$j(F)_z = \pi a(F)_z b(F)_z, \quad l(F'_z(C_1)) = \pi(a(F)_z + b(F)_z) \le 2\pi a(F)_z,$$

where $b(F)_z = \min_{v\in S^1} \|F'_z(v)\|$, one easily derives the following

LEMMA 6.3. (a) *There exist* $\tau_1, \tau_2 > 0$ *such that for any* $z \in \mathbb{R}^2$ *with* $c(F)_z \neq 0$, *and* $\varepsilon > 0$ *small enough* (*depending on* z) *the following inequality holds*:

$$\tau_1 |c(F)_z| \left(\frac{j(F)_z}{\mathrm{e.r.}(F)_z}\right)^3 \varepsilon^3 - \tau_2 D^2(F)_z a(F)_z \varepsilon^3 \le \mathrm{vol}(F(D_{z,\varepsilon}), D^{\mathrm{curv}}_{F,z}).$$

(b) *There exists* $\tau_3 > 0$ *such that for any* $z \in \mathbb{R}^2$, *any smooth curve* $P \subset D_{z,\varepsilon}$ *which is a topological diameter of* $D_{z,\varepsilon}$, *and any* $\varepsilon > 0$ *small enough* (*depending on* z) *the following inequality holds*:

$$\begin{aligned}|(\mathrm{vol}(F(D')) - \mathrm{vol}(F(D''))) - j(F)_z(\mathrm{vol}(D') - \mathrm{vol}(D''))|& \\ \le \tau_3 D^2(F)_z a(F)_z \varepsilon^2 (2\pi\varepsilon + l(P)),&\end{aligned}$$

where D' *and* D'' *denote the two parts of* $D_{z,\varepsilon}$, *separated by* P.

Proof of Theorem 6.1. Assume F is not injective. Fix $\{c_i\}_{i\in\mathbb{N}}$, $\{z_i\}_{i\in\mathbb{N}}$ and $\varrho > 0$ as in Corollary 5.2. After scaling we can achieve $\varrho = 1$.

Let $\varepsilon>0$ be small. We put $D^+_{z_i,\varepsilon} = D_{z_i,\varepsilon} \cap D_{\|z_i\|}$ and $D^-_{z_i,\varepsilon} = D_{z_i,\varepsilon} \setminus D_{\|z_i\|}$.

The first observation is that $\mathrm{vol}(D^+_{z_i,\varepsilon}) - \mathrm{vol}(D^-_{z_i,\varepsilon}) < 0$ for all i and all small $\varepsilon > 0$. Therefore $j(F)_{z_i}(\mathrm{vol}(D^+_{z_i,\varepsilon}) - \mathrm{vol}(D^-_{z_i,\varepsilon})) < 0$.

Let $P_{i,\varepsilon}$ denote the arc of $C_{\|z_i\|}$ inside $D_{z_i,\varepsilon}$. Then $l(P_{i,\varepsilon}) < 3\varepsilon$ for $\varepsilon > 0$ small. It follows from Lemma 6.3(b) and the assumptions of the theorem that

$$\mathrm{vol}(F(D^+_{z_i,\varepsilon})) - \mathrm{vol}(F(D^-_{z_i,\varepsilon})) < \tau_3 D^2(F)_{z_i} a(F)_{z_i} \varepsilon^2 (2\pi\varepsilon + 3\varepsilon)$$

and

$$\tau_3 D^2(F)_{z_i} a(F)_{z_i} \varepsilon^2 (2\pi\varepsilon + 3\varepsilon) \le (2\pi + 3)\tau_3 \kappa_2 \|z_i\|^{3\lambda+3} \varepsilon^3$$

whenever i is large and $\varepsilon > 0$ is small enough (depending on i). Since

$$\mathrm{vol}(F(D_{z_i,\varepsilon}), D^{\mathrm{curv}}_{F,z_i}) \sim \mathrm{vol}(F(D^+_{z_i,\varepsilon})) - \mathrm{vol}(F(D^-_{z_i,\varepsilon}))$$

as $\varepsilon \to 0$, we have

$$\mathrm{vol}(F(D_{z_i,\varepsilon}), D^{\mathrm{curv}}_{F,z_i}) < 2(\mathrm{vol}(F(D^+_{z_i,\varepsilon})) - \mathrm{vol}(F(D^-_{z_i,\varepsilon})))$$

for $\varepsilon > 0$ small (depending on i). On the other hand, by Lemma 6.3(a),

$$\tau_1 c_i^3 j(F)^3_{z_i} \|z_i\|^3 \varepsilon^3 - \tau_2 D^2(F)_z a(F)_{z_i} \varepsilon^3 \le \mathrm{vol}(F(D_{z_i,\varepsilon}), D^{\mathrm{curv}}_{z_i})$$

for i large and $\varepsilon > 0$ small enough (depending on i). By the assumptions

$$\|z_i\|^{3\lambda+3} \varepsilon^3 (\tau_1 \kappa_1^3 c_i^3 - \tau_2 \kappa_2) \le \tau_1 c_i^3 j(F)^3_{z_i} \|z_i\|^3 \varepsilon^3 - \tau_2 D^2(F)_z a(F)_{z_i} \varepsilon^3.$$

Summing up these 5 inequalities (and using the equivalence above) we get

$$\tau_1 \kappa_1^3 c_i^3 - \tau_2 \kappa_2 < 2(2\pi + 3)\tau_3 \kappa_2,$$

which is obviously violated for i large, a contradiction. ∎

References

[BR] A. Białynicki-Birula and M. Rosenlicht, *Injective morphisms of real algebraic varieties*, Proc. Amer. Math. Soc. 13 (1962), 200–204.

[F] W. Fulton, *Algebraic Topology (a First Course)*, Grad. Texts in Math. 153, Springer, 1995.

[Gw] J. Gwoździewicz, *The Real Jacobian Conjecture for polynomials of degree 3*, preprint, 1999.

[Ha] J. Hadamard, *Sur les transformations ponctuelles*, Bull. Soc. Math. France 34 (1906), 71–84.

[KR] K. Kurdyka and K. Rusek, *Polynomial rational bijections of* $\mathbb{R}^n$, Proc. Amer. Math. Soc. 112 (1988), 804–808.

[P] S. Pinchuk, *A counterexample to the strong real Jacobian conjecture*, Math. Z. 217 (1994), 1–4.

A. Razmadze Mathematical Institute
Georgian Academy of Sciences
Alexidze Str. 1
Tbilisi 380093, Republic of Georgia
E-mail: gubel@rmi.acnet.ge

Current address:
FB Mathematik/Informatik
Universität Osnabrück
49069 Osnabrück, Germany
E-mail: jgubelad@mathematik.uni-osnabrueck.de

Reçu par la Rédaction le 5.2.1999
Révisé le 31.5.2000