On the problem of solving $F(x) = y$ in the plane

I.P. Malta, N.C. Saldanha and C. Tomei
Departamento de Matemática, PUC-Rio [1]

Abstract: The knowledge of a critical set together with its image provides a wealth of information about the global geometry of mappings from the plane to the plane. In this note, we describe some properties satisfied by the critical sets and indicate how they can be useful in the explicit construction of the critical sets and in the solution of $F(x) = y$.

In this note[2], we describe a theoretical setup to study the problem of solving $F(x) = y$ for a large class of mappings F from the plane to the plane. The motivation to consider this problem came from the one dimensional Riemann problem, that is, a conservation law

$$U_t + f(U)_x = 0, \qquad f : \mathbf{R}^2 \to \mathbf{R}^2$$

with initial condition

$$U(0,x) = \begin{cases} U_0 & \text{for } x < 0, \\ U_1 & \text{for } x > 0. \end{cases}$$

An elementary shock-wave solution of the Riemann problem is given by

$$U(t,x) = \begin{cases} U_0 & \text{for } x - st < 0, \\ U_1 & \text{for } x - st > 0. \end{cases}$$

whenever the Rankine-Hugoniot condition

$$f(U_1) - f(U_0) = s(U_1 - U_0)$$

holds. The search for elementary shock-wave solutions then leads us to the following question: for given $s \in \mathbf{R}$ and $U_0 \in \mathbf{R}^2$, how many values of $U_1 \in \mathbf{R}^2$ are there so that (s, U_0, U_1) gives rise to an elementary shock-wave solution, and how does this number vary as (s, U_0) vary?

Rewriting the Rankine-Hugoniot equation as

$$f(U_1) - sU_1 = f(U_0) - sU_0$$

and setting $F_s = f - sI$, one sees that (s, U_0, U_1) gives rise to an elementary shock-wave solution of the Riemann problem if and only if U_1 is a solution of the non-linear equation

$$F_s(U) = F_s(U_0).$$

[1] Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, RJ 22453, Brasil
[2] Work supported by SCT and CNPq, Brazil

Thus the above question naturally induces the following problem.

Problem 1: Given $F : \mathbf{R}^2 \to \mathbf{R}^2$ and $y \in \mathbf{R}^2$, how many solutions does the equation $F(x) = y$ have, and how does the number of solutions vary with y?

The theoretical considerations in this note are also being used in the construction of a program to solve the equation $F(x) = y$ numerically, for a rather general class of mappings F described below. The key point is that the topological arguments not only provide a satisfactory counting of the solutions of the equation, but also show us how to implement routines which obtain initial conditions for local iterations (like Newton's method) which indeed converge to the desired solutions. Details of the program will be described elsewhere [MST].

In [MT] it is proved that, for F in a suitable class of C^∞ "nice" mappings of the plane into itself, all the information needed to answer Problem 1 is given by the action of F on the set of critical points. Nice mappings will be defined later; over compact domains, nice mappings are generic. Rather surprisingly, not any set of curves in the plane, together with a putative set of image curves, can be the critical set of a nice mapping in the plane. Some obvious restrictions arise from considerations of differential topology. In the seventies, Blank and Troyer ([B],[T]) obtained necessary and sufficient conditions for a set of curves to be the image of the boundary under an immersion of an n-holed disk in the plane. From their work, we derive a satisfactory answer to the following problem, which appears naturally in the computation of the critical set of a given nice mapping F.

Problem 2: Given a set C' of critical points of F, is there a nice mapping G such that G agrees with F in a neighborhood of C' and C' is the whole critical set of G?

At the end of this note, we make some comments about a delicate issue in the project: given a nice mapping F and some curves which are shown, by making use of the appropriate tests, to be *all* the critical curves of some nice mapping G, how can we be sure that there are no other critical curves?

In this text, we provide no proofs (to be found in the references) and frequently avoid a complicated (and precise) description of the results by making use of examples, which are supposed to convey the spirit of the techniques employed. We would like to thank the referee for his comments, indicative of a careful reading.

We begin by recalling some standard definitions. A mapping F from the plane to itself is said to be proper if the inverse image of any compact set in the plane is compact. Clearly, $F : \mathbf{R}^2 \to \mathbf{R}^2$ is proper if and only if the mapping goes to infinity at infinity, and so F can be extended continuously from the Riemann sphere to itself by defining the value of the extension $\tilde{F}$ at infinity to be infinity. Continuous proper mappings F from the plane to itself have a topological degree, which coincides with the degree of the extension $\tilde{F}$ [M]. A point in the domain of F is regular if the differential DF at this point is invertible. Points which are not regular are called critical, and their images are the critical values of F. The set of critical points, denoted by $C(F)$ or simply C, and its image under F are called the critical sets of F. Points which are not critical values are called regular values.

To explain how all the information we need to solve the equation $F(x) = y$, for F in a suitable class of C^∞ mappings, is given by the critical sets of F, we consider the mapping $F : \mathbf{R}^2 \to \mathbf{R}^2$ given by

$$F(u, v) = (-6u^4 - 6u^2v^2 + uv^3 + 6v^4 - u, \frac{25}{24}u^4 + u^3v + u^2v^2 + \frac{1}{6}uv^3 - v^4 - v).$$

This mapping is proper and its critical sets C and $F(C)$ are given in Figure 1, while the pre-image $F^{-1}(F(C))$ is given in Figure 2. We use the same letter to denote a set and its image under F.

Now, if we denote by T_∞ the unbounded component of $\mathbf{R}^2 - F(C)$ and by T_0 and T_1 the bounded ones, we will have, from the properness of F, that F restricted to any connected component of $F^{-1}(T_\alpha), \alpha = 0, 1, \infty$, is a covering mapping and the action of F is described in Figure 2. So, in particular, F restricted to any connected component of $F^{-1}(T_i), i = 0, 1$, is a diffeomorphism onto T_i and $F^{-1}(T_\infty)$ covers T_∞ twice. Hence the equation $F(x) = y$ has 2, 4 or 6 solutions depending on whether y belongs to T_∞, T_0 or T_1 respectively. Indeed, as proved in [MT], for a nice F, the number of solutions of $F(x) = y$, for y varying in the plane, can be obtained simply from knowledge of C and $F(C)$. The additional knowledge of $F^{-1}(F(C))$ gives us information about where to look for initial conditions for a possible iterative method to solve the equation numerically.

In order to state the hypothesis on F, we recall the concept of Whitney singularities [W]. A fold point of F (resp., cusp point) is a critical point x for which there are local orientation preserving diffeomorphisms around x and $F(x)$ onto neighborhoods of the origin of the plane in which F takes the form (1) (resp., (2)) below.

$$F(u, v) = (u, v^2) \tag{1}$$

$$F(u, v) = (u, av^3 - uv), \qquad a = \pm 1 \tag{2}$$

By a celebrated theorem of Whitney [W], generically in an appropriate topology in the space of smooth mappings from the plane to itself, critical points are either folds or cusps. Mappings satisfying the three conditions below will be called nice.

(a) F is a smooth proper mapping from the plane to itself.

(b) $C = C(F)$ is bounded and each critical point is a fold or a cusp point.

(c) Images of critical curves may only intersect at a finite number of points, and the pre-image of such a point meets $C(F)$ at exactly two fold points.

With the techniques employed in [W], one shows that nice mappings are generic in the class of smooth proper mappings with bounded critical set (in a suitable C^r topology). The requirement that each critical point is a Whitney singularity implies that zero is a regular value of $\det DF$. Thus, from condition (b), the set of critical points C is a finite disjoint union of simple closed curves and there is only a finite number of cusp points. Condition (a) implies that F has a topological degree $d = \deg F$. This degree is given by the number of solutions x_j of the equation $F(x) = y$ for any regular value y, counted with the sign of $\det DF(x_j)$ [M]. Since by (b) the set of critical points C is compact, we have that $\mathbf{R}^2 - F(C)$ has exactly one unbounded connected component T_∞. Clearly, $F^{-1}(T_\infty)$ is the unbounded connected component of $\mathbf{R}^2 - F^{-1}(F(C))$, so the local orientation of F at any solution of $F(x) = y$ for y in T_∞ has to be the same. Thus, the number of solutions of this equation for y in T_∞, which is positive, equals the absolute value of the topological degree of F. In particular, $d \neq 0$.

From (b), the restriction of F to any critical curve Γ_i , $F(\Gamma_i)$, is a continuous locally injective curve and we can define the turning number, $\tau(F|_{\Gamma_i})$ of $F(\Gamma_i)$ as the Brouwer degree of the mapping

$$\vartheta \mapsto \frac{F(\gamma(\vartheta + \delta)) - F(\gamma(\vartheta))}{|F(\gamma(\vartheta + \delta)) - F(\gamma(\vartheta))|},$$

where $\gamma : S^1 \to \mathbf{R}^2$ is a regular orientation preserving parametrization of Γ_i and δ is any small positive angle for which $F|_{[\gamma(\vartheta),\gamma(\vartheta+\delta)]}$ is injective for all $\vartheta \in S^1$. Condition (c) implies that we can compute the turning number $\tau(F|_{\Gamma_i})$ from the Seifert circles of $F(\Gamma_i)$ as follows. Consider the orientation induced in the curve $F(\Gamma_i)$ by the positive (counterclockwise) orientation of Γ_i . Now change slightly the curve $F(\Gamma_i)$ near self-intersection points as in Figure 3. Then $F(\Gamma_i)$ splits into a disjoint union of oriented simple curves – the turning number is the number of such simple curves counted with a sign defined by their orientation.

Since critical points are fold or cusp points, we can orient the image $F(\Gamma_i)$ of each critical curve Γ_i so that a small disk around any fold point in Γ_i is sent by F to the left side of the oriented curve $F(\Gamma_i)$. In particular, each image $F(x)$ of a cusp point $x \in \Gamma_i$ points to the right side of the oriented curve $F(\Gamma_i)$ (see Figure 4). This orientation will be said to be given by the sense of folding. The turning $w(\Gamma_i)$ of the curve $F(\Gamma_i)$ oriented by the sense of folding is defined by $w(\Gamma_i) = \pm\tau(F|_{\Gamma_i})$. This sign is positive if this orientation agrees with the one induced by F on $F(\Gamma_i)$ from the positive orientation on Γ_i .

The three results below are stringent relations among the numbers defined so far (see [MT] for proofs).

Theorem A: Let F be a nice mapping. Then

$$|\mathrm{deg} F| = k - 2w + 1 > 0,$$

where k is the number of cusp points and $w = \sum_{\Gamma_i \subset C} w(\Gamma_i)$.

Let $F : \mathbf{R}^2 \to \mathbf{R}^2$ be a nice mapping and $C' = \bigcup_{i=1}^n \Gamma_i$ a union of critical curves of F. Let Γ_0 be a circle around C' and suppose that $\Gamma_0 \cap C(F) = \emptyset$. Let D_0 be the open disk bounded by Γ_0 . Let $w(\Gamma_0)$ be the turning number of $F(\Gamma_0)$ oriented so that a small neighborhood in $\overline{D}_0$ of a point in Γ_0 is sent by F to the left side of $F(\Gamma_0)$, and $w(\Gamma_i)$ for $i = 1, \ldots, n$ be the turning of $F(\Gamma_i)$ oriented by the sense of folding.

Proposition [MT]**:** If $D_0 \cap C(F) = C'$ then

$$w(\Gamma_0) = k - 2w(C') + 1,$$

where $w(C') = \sum_{i=1}^{n} w(\Gamma_i)$ and k is the number of cusp points in C'.

Corollary: If $C' = C(F)$ then $|\mathrm{deg} F| = w(\Gamma_0)$.

We show by an example how to solve Problem 2. Suppose we have found a critical curve Γ_1 of a nice mapping F such that the image curve oriented by the sense of folding and the behavior of F in a neighborhood of Γ_1 are given in Figures 5 and 6 respectively.

Since $w(\Gamma_1) = 1$ and Γ_1 has one cusp point, we conclude from Theorem A that Γ_1 cannot be the whole critical set of F, so we have to look for more critical curves.

Suppose we find another critical curve Γ_2 as in Figure 7, close to which the local behavior of F is given in Figure 8.

From Theorem A we learn that if $\Gamma_1 \cup \Gamma_2$ is a critical set of a nice mapping G then $|\deg G| = 1$. So, if we take a circle around the curves in the range, its pre-image is a simple, closed regular curve Γ_0 around the critical curves Γ_1 and Γ_2 as in Figure 9. By the proposition and corollary above, $w(\Gamma_0) = 1$. The orientation in $F(\Gamma_0)$ means that any point in the open disk bounded by Γ_0 is sent by F to the left side of $F(\Gamma_0)$.

Clearly the problem of existence of a nice mapping G as required is then reduced to the following question. Is there an extension G of $F|_{\Gamma_0 \cup \Gamma_1 \cup \Gamma_2}$ to the closed disk $\overline{D_0}$ bounded by Γ_0 such that G is an immersion outside the curves Γ_1 and Γ_2 and preserves the behavior of F near $\Gamma_1 \cup \Gamma_2$? To answer this question, first observe that this problem decouples in similar problems. Let $C' = \Gamma_0 \cup \Gamma_1 \cup \Gamma_2$ and $g = F|_{C'}$.

(i) For $i = 1, 2$, can $g|_{\Gamma_i}$ be extended to the closed disk $\overline{D_i}$ bounded by Γ_i , as an immersion in D_i and such that the extension preserves the behavior of $F|_{\overline{D_i}}$ near Γ_i ?

(ii) Can g be extended to $A = \overline{D_0} - (D_1 \cup D_2)$ as an immersion outside of the boundary of A preserving the behavior of $F|_A$ near ∂A?

Clearly, for $i = 1, 2$, $g|_{\Gamma_i}$ has an extension as a diffeomorphism from D_i onto the open disk bounded by $F(\Gamma_i)$. To give an answer to (ii), we use a slight adaptation of a criterion introduced by Blank [B] and Troyer [T]. To explain how to do this, we consider first a simpler problem. Suppose that Γ is a critical curve of a nice mapping H and that $h = H|_\Gamma$ acts as in Figure 10.

The problem is to decide if there exists an extension of h to the closed disk $\overline{D}$ bounded by Γ as an immersion in D, having x and y as cusp points, and such that a small neighborhood in $\overline{D}$ of a point in $\Gamma - \{x, y\}$ is sent to the left side of the oriented curve $h(\Gamma)$. Proceed as follows. Let T_∞ be the unbounded connected component of $\mathbf{R}^2 - h(\Gamma)$. Choose a point p_i in each bounded connected component of $\mathbf{R}^2 - h(\Gamma)$. Let r_i be a proper embedding of the non-negative real axis into the plane with $r_i(0) = p_i$ (called a ray from p_i) such that the rays $r_1, \ldots, r_m$ do not contain intersection points of $h(\Gamma)$, cut $h(\Gamma)$ transverswally and are pairwise disjoint. Now assign to each intersection of a ray with the curve a positive or a negative sign, depending on whether the curve crosses the ray from right to left (that is, p_i is on the left side of the curve) or the curve crosses the ray from left to right. Complete the set of rays by choosing for each cusp point x a ray from $h(x)$ assigning to this point a negative sign (see Figure 11).

Now construct a word (the Blank word for the oriented curve $h(\Gamma)$) by following the orientation of the curve $h(\Gamma)$ and collecting the intersection points with the rays, keeping track of signs and rays. Clearly the word is defined up to cyclic permutation. In this example, a Blank word is given by

$$BW = r_1^+ r_2^+ r_3^+ r_4^+ r_5^+ r_6^+ r_7^+ r_3^+ r_4^+ r_5^+ r_2^+ r_3^+ r_5^- r_6^+ r_7^- .$$

We say that a Blank word admits a simplification if there exists a pair r_i^+, r_i^- such that (after a cyclic permutation if necessary) there are no letters with negative exponent between r_i^+ and r_i^- . If this is the case, a simplified Blank word is obtained by eliminating the subword $r_i^+ \ldots r_i^-$ (or $r_i^- \ldots r_i^+$). We say that a Blank word groups (or has a grouping) if there are successive simplifications such that the final Blank word has no letters with negative exponents. In the example, a possible choice of simplification is

$$BW \to r_1^+ r_2^+ r_3^+ r_5^+ r_6^+ r_7^+ r_3^+ r_4^+ r_6^+ r_7^- \to r_1^+ r_2^+ r_3^+ r_5^+ r_6^+$$

and so the Blank word of the example groups. The fact that the Blank word groups, together with the fact that the turning number minus the number of cusps equals one, is sufficient (and necessary!) to guarantee the existence of the desired extension. The complete description of Blank's criterion can be found in [B], while the mild extension being used is in [MST].

Returning to the example given in Figure 9, we follow (a slight modification of) a procedure due to Troyer ([T]) which is the analogue of Blank's criterion to the case where the domain of the extension being sought consists of more than one curve. We begin by constructing rays as before (see Figure 12), and collecting one word for each curve $g(\Gamma_i), i = 0, 1, 2$. A Blank word for $g(C')$ is a single word obtained as the adjunction of a cyclic permutation of each word. In the example, denoting by $BW(\Gamma_i)$ the Blank word of $g(\Gamma_i), i = 0, 1, 2$, we have

$$BW(\Gamma_0) = a^+ b^+ c^+ d^+ e^+ f^+ q^+ r^+,$$

$$BW(\Gamma_1) = d^+ e^+ f^+ c^-,$$

$$BW(\Gamma_2) = c^+ d^+ f^- q^+ r^- a^-,$$

so a possible Blank word for $g(\Gamma_0 \cup \Gamma_1 \cup \Gamma_2)$ is

$$BW(\Gamma_0 \cup \Gamma_1 \cup \Gamma_2) = d^+ e^+ f^+ c^- c^+ d^+ f^- q^+ r^- a^- a^+ b^+ c^+ d^+ e^+ f^+ q^+ r^+$$

which has a grouping. As before, since $w(\Gamma_0) + w(\Gamma_1) + w(\Gamma_2) -$ (number of cusp points) $= -1$, and a Blank word groups, we can guarantee the existence of the desired extension.

Let $A_1, \ldots, A_m$ be the connected components of $D_0 - C'$, with the notation defined above. Then for each $p = 1, \ldots, m$, the boundary of A_p is $\partial A_p = \Gamma_{p_0} \cup (\bigcup_{i=1}^{n_p} \Gamma_{p_j})$ where Γ_{p_0} is the exterior boundary of A_p, that is, A_p is contained in the disk bounded by Γ_{p_0}. As in the example, the problem of existence of a nice mapping G which agrees with F in C', has the same behavior as F in a neighborhood of C' and such that $C(G) = C'$ can be reduced to the existence of extensions of $F|\partial A_p$ to A_p as an immersion in A_p preserving the behavior of F in a neighborhood of ∂A_p in A_p. So we have to make a distinction among the cusps in ∂A_p in such a way as to select the ones which are detected by the action of F in A_p. Let D_i be the open disk bounded by the critical curve Γ_i . A cusp point $x \in \Gamma_i$ will be called inward if, for each sufficiently small neighborhood U of x, F maps $U \cap \overline{D}_i$ onto a neighborhood of $F(x)$. It is clear from the local form (2) that if a cusp point x is not inward, then F maps $(\mathbf{R}^2 - D_i) \cap U$ onto a neighborhood of $F(x)$. In this case, x is said to be an outward cusp point. In other words, the fact that x is a cusp point is detected by the action of F in D_i or in $\mathbf{R}^2 - \overline{D}_i$ if x is an inward or outward cusp respectively. These cases correspond to Figures 13 (a) and (b).

Clearly, for the problem of existence of the desired extension of $F|\partial A_p$, we have to consider as cusp points only the inward cusp points of Γ_{p_0}, and the outward cusp points of $\Gamma_{p_j}, j = 1, \ldots, n_p$. So, for each p, construct a set of rays for the curves $F(\partial A_p)$ oriented by the sense of folding considering only the cusps described above. In Figures 14 (a) and (b), we draw the rays associated to the components A_1 and A_2 for F given in Figure 1.

For each p, construct the Blank word of $F(\Gamma_{p_j}), j = 0, \ldots, n_p$ and consider the set of Blank words for $F(\partial A_p)$. For each p, let $k_p = k_{*p_0} + \sum_{j=1}^{n_p} k^*_{p_j}$, where k_{*p_0} and $k^*_{p_j}$ are, respectively, the number of inward and outward cusps in the corresponding curve.

Theorem B: ([MST]) C' is the set of critical points of a nice mapping G which agrees with F in a neighborhood of C' if and only if

(i) $k_p - (\sum_{j=0}^{n_p} w(\Gamma_{p_j})) - n_p + 1 = 0$ for $p = 1, \ldots, m$, and

(ii) for each $p = 1, \ldots, m$, there exists a Blank word for $F(\partial A_p)$ which has a grouping.

Moreover, if (i) and (ii) are satisfied, then $|\deg G| = w(\Gamma_0)$.

Remark: Condition (i) simply says that for each p, $\chi(\overline{A}_p) = w(\partial A_p) - k_p$, where $\chi(\overline{A}_p)$ is the Euler characteristic of $\overline{A}_p$, which is $1 - n_p$.

We are still left with the following intriguing issue: once we found a (sub)set of critical curves of the mapping F which is indeed the critical set of a mapping G (since it satisfies the tests described above), how can we be sure that F does not have other critical curves? The answer is obvious: this cannot be guaranteed by mere topological arguments.

The problem is already present in the one dimensional case in two different versions: how can we be sure, given a mapping from the line to itself that we know all its zeros, or all the zeros of its derivative? There are two obvious difficulties: a computer program will search for zeros in a bounded set and not beyond a certain level of refinement within this set. Implicitly, we assume the knowledge of a priori estimates which guarantee that there are no zeros outside the set being scanned and that, within this set, the mapping does not oscillate so much as to generate additional zeros. Thus, unless the program itself generates the required estimates, it cannot certify that all zeros will be found.

We provide two classes of examples in which the set of critical curves is not found completely. The reader will have no trouble in convincing himself that the examples are typical. In Figure 15, the knowledge of the critical curve Γ together with the behavior of the mapping F at infinity (its degree) is not enough to decide whether the critical set of F is (a) or (b). The problem is that one can draw a curve γ around the remaining critical curves in (b) for which $F(\gamma)$ is a simple closed curve: if you never enter the disk D bounded by γ you will never distinguish between the complicated behavior of F in D described in (b) and the simple one (a diffeomorphism) shown in (a). So, unless the original search for critical curves spontaneously browsed through D, the topological arguments presented in this note would not indicate that D should be searched for critical points.

The other example is shown in Figure 16. Knowing curve Γ and the degree of F would not induce the program to search for the (possibly missing) two critical curves. The reason is that the image of a simple curve tightly surrounding Γ has the same topological behavior as the image of a large circle: both are taken to (isotopic) curves of equal turnings. So, again, from the topological information, both situations are undistinguishable.

Bibliography

[B] Blank, S.J.: *Extending immersions of the circle*, Exposé 342, Séminaire Bourbaki 1967-68, Benjamin, NY, 1969.

[M] Milnor, J.W.: *Topology from the differentiable viewpoint*, University Press of Virginia, 1969.

[MST] Malta, I., Saldanha, N.C., Tomei, C.: *Critical sets and the numerical inversion of mappings in the plane*, in preparation.

[MT] Malta, I., Tomei, C.: *Singularities of vector fields arising from one dimensional Riemann problems*, J.Diff.Eq., 93 (1991).

[T] Troyer, S.F.: *Extending a boundary immersion to the disk with n holes*, Ph. D. Dissertation, Northeastern U., Boston, Mass., 1973.

[W] Whitney, H.: *On singularities of mappings of Euclidean spaces, I: mappings of the plane into the plane*, Ann. of Math. (1955), 374-410.

malta@mat.puc-rio.br
nicolau@mat.puc-rio.br, http://www.mat.puc-rio.br/∼nicolau/
tomei@mat.puc-rio.br