On the Zeros of Harmonic Mappings: Analysis, Computation and Application

vorgelegt von

M. Sc.

Jan Zur

ORCID: 0000-0002-5611-8900

an der Fakultät II – Mathematik und Naturwissenschaften –
der Technischen Universität Berlin
zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften
— Dr. rer. nat. —
genehmigte Dissertation

Promotionsausschuss:

Vorsitzender:	Prof. Dr. Stefan Felsner (TU Berlin)
Gutachter:	Prof. Dr. Jörg Liesen (TU Berlin)
Gutachter:	Prof. Dr. Yuji Nakatsukasa (University of Oxford)
Gutachter:	Prof. Dr. Elias Wegert (TU Bergakademie Freiberg)

Tag der wissenschaftlichen Aussprache: 23. Juni 2022

Berlin 2022

Abstract

In this doctoral thesis, we consider the zeros of harmonic mappings in the complex plane. Our study was originally motivated by the theory of gravitational lensing in astrophysics, where special cases of such functions and their zeros play an important role. However, in this work we focus on much more general functions. Our results range from purely analytical to numerical aspects. We sharpen bounds previously published by other authors on the number of zeros of rational harmonic functions $f(z) = r(z) - \overline{z}$ and harmonic polynomials $f(z) = p(z) + \overline{q(z)}$. Moreover, we develop the first numerical continuation method that is guaranteed to find all zeros of a (non-singular and non-degenerate) harmonic mapping f. Our method works without any prior knowledge about the number or the position of the zeros. Numerical experiments indicate that our method is extremely fast and highly accurate. The foundation of both our analytical and our numerical results is an extensive study of how the number of solutions of $f(z) = \eta$ changes as the parameter $\eta \in \mathbb{C}$ varies. For this purpose, we use mathematical tools from very different areas, such as the argument principle for harmonic mappings or convergence results of Newton's method in Banach spaces. As it turns out, the position of η with respect to the caustics of f is crucial for determining the number of solutions of $f(z) = \eta$. We formalize this observation by providing a complete theory that is of its own interest, and a starting point to further advance the understanding of harmonic mappings in the future.

Acknowledgments

First and foremost, I wish to express the deepest gratitude to my supervisor Jörg Liesen for his guidance and encouragement throughout my dissertation project. He was always willing and interested to support me in any conceivable way. The stimulating and pleasant working atmosphere he created in his research group made my time there a great experience and undoubtedly contributed significantly to this work.

I am very grateful that in Olivier Sète I had a collaborator and friend who shared my enthusiasm for the theory of harmonic mappings. Without our joint work, this thesis would be a very different one.

Finally, I would like to thank my family: Johanna and Jakob, for cheering me up when I got stuck, and my wonderful wife Kristin for her endless love and patience over the years.

Contents

1 Introduction

This cumulative doctoral thesis in mathematics is devoted to the study of the zeros of harmonic mappings in the complex plane. It covers analytical as well as numerical aspects, and hence lies within the intersection of complex analysis and numerical analysis. The presented results have been previously published in the following peer-reviewed journal articles:

[44] Jörg Liesen and Jan Zur, *How constant shifts affect the zeros of certain rational harmonic functions*, Comput. Methods Funct. Theory, 18(4) (2018), pp. 583–607, doi:10.1007/s40315-018-0240-8.

[45] Jörg Liesen and Jan Zur, *The maximum number of zeros of $r(z)-\overline{z}$ revisited*, Comput. Methods Funct. Theory, 18(3) (2018), pp. 463–472, doi:10.1007/s40315-017-0231-1.

[72] Olivier Sète and Jan Zur, *A Newton method for harmonic mappings in the plane*, IMA J. Numer. Anal., 40(4) (2020), pp. 2777–2801, doi:10.1093/imanum/drz042.

[73] Olivier Sète and Jan Zur, *Number and location of pre-images under harmonic mappings in the plane*, Ann. Fenn. Math., 46(1) (2021), pp. 225–247, doi:10.5186/aasfm.2021.4614.

[74] Olivier Sète and Jan Zur, *The transport of images method: computing all zeros of harmonic mappings by continuation*, IMA J. Numer. Anal., (2021), doi:10.1093/imanum/drab040. Online first publication.

The main body of this thesis links the above publications, provides an introduction to the respective research topics, and summarizes the main results of the articles. Preprint versions of the articles are included as Appendices A, B, C, D and E.

1.1 Zeros of harmonic mappings in the complex plane

Computing the zeros of a given function is one of the core problems in mathematics. It appears throughout all areas, as well as all levels of difficulty. From the quadratic formula for the zeros of second degree polynomials, which is familiar to all high school students, to the Riemann zeta function, where proving that all nontrivial zeros have real part $\frac{1}{2}$ would answer one of the most important open problems in modern mathematics.

Also this thesis deals with the analysis and computation of zeros. Our functions of interest are *harmonic mappings*. A harmonic mapping f is a complex-valued function, defined on an open subset Ω of the complex plane, which satisfies Laplace's equation, i.e.,

$$f : \Omega \subseteq \mathbb{C} \to \mathbb{C} \quad \text{with} \quad \Delta f = 0. \tag{1.1}$$

According to Duren [24, Sect. 1.6], the modern treatment of harmonic mappings began in 1984 with the seminal paper of Clunie and Sheil-Small [19], which drew the attention of many complex analysts to these functions. Since then, the theory of harmonic mappings became an active research topic; see, e.g., the large collection of open problems by Bshouty and Lyzzaik [17] from 2010.

Harmonic mappings have a local decomposition

$$f(z) = h(z) + \overline{g(z)}, \tag{1.2}$$

with analytic functions h and g, which are unique up to an additive constant [24, p. 7], but harmonic mappings are *not analytic* in general. Unfortunately, many properties of analytic functions do not hold for harmonic mappings. For example, the composition of two harmonic mappings is not necessarily a harmonic mapping, while the composition of two analytic functions is always analytic. More generally, the composition $\varphi \circ f$ of an analytic function φ and a harmonic mapping f is not harmonic in general.

Another difference is that a zero z_0 of a harmonic mapping f cannot be factored out in general, i.e., there exists no harmonic mapping g with $f(z) = (z - z_0)g(z)$ or with $f(z) = \overline{(z - z_0)}g(z)$; see [71, p. 13]. Therefore, zeros of harmonic mappings do not have multiplicities in the usual sense and by *number of zeros* we refer to the number of distinct complex points at which f vanishes. However, using the Poincaré index, it is still possible to count zeros of harmonic mappings via the argument principle; see Section 2.1. Also, the local decomposition (1.2) provides a powerful tool to study harmonic mappings, which we frequently used in our publications.

Given a harmonic mapping f and a complex number $\eta \in \mathbb{C}$, the recurring questions of this work are:

Q1: How does η affect the zeros of the shifted function $f_\eta(z) = f(z) - \eta$?

Q2: How do the pre-images of η under f change with η?

Q3: How do the solutions of $f(z) = \eta$ depend on η?

These questions are all quite similar, and it is a matter of taste which one we use. Depending on our perspective we prefer one or the other for the simplicity of our presentation. We continue with the third formulation.

At first glance, it may seem cumbersome to study the equation $f(z) = \eta$ for the general parameter $\eta \in \mathbb{C}$, although we are only interested in the zeros of f, i.e., in the case $\eta = 0$. But this more general problem is essential for the analytical study of the zeros of harmonic mappings and it has also important practical applications; see Section 1.2. Obviously, the positions of the solutions of $f(z) = \eta$ depend on η, but, somewhat surprisingly, also the *number of solutions* of $f(z) = \eta$ depends on η. As an example, consider the harmonic polynomial $f(z) = z^2 - \overline{z}$, which is one of the simplest harmonic

mappings one can think of. A direct calculation shows that $f(z) = 0$ has the four solutions 0, 1, $e^{i\frac{2\pi}{3}}$ and $e^{i\frac{4\pi}{3}}$, while $f(z) = 2$ has only the two solutions -1 and 2. This property makes the study of the solutions of $f(z) = \eta$ truly fascinating. If the numbers of solutions of $f(z) = \eta$ and $f(z) = \xi$ differ and are finite, then there must exist some kind of "singular point" on each path from η to ξ where the number of solutions jumps from one integer number to another one. Analyzing *when*, *where* and *how* this happens provides remarkable insights into the nature of harmonic mappings. It allows to count and locate the solutions of $f(z) = \eta$, and thus also the zeros of f. In the end, our analysis will lead to a numerical method to compute *all* zeros of certain harmonic mappings.

1.2 Application to gravitational lensing

One particular application of harmonic mappings and their zeros is *gravitational lensing*. This astrophysical phenomenon occurs when light of a far away source is deflected by massive objects that form a *gravitational lens*. The effect on the observed image ranges from magnification (microlensing) through weak distortions (weak lensing) to strong distortions or even multiple images (strong lensing); see, e.g., [20, Sect. 1.4] and [66, p. 47].

Gravitational lensing was first correctly described by Einstein in 1915/16 using his general theory of relativity. But also long before that, it was suspected that gravity influences the light, e.g., by Newton. In 1919, Eddington confirmed Einstein's prediction by observations during a total solar eclipse. This was also the first confirmation of general relativity; see [66, Ch. 1] and [69, Sect. 1.1] for further historical comments.

Starting from the 1970s, when the first gravitational lenses were actually discovered, gravitational lensing has proven to be a very powerful tool for studying astronomical objects. Gravitational lenses magnify the light and allow us to observe extremely distant objects. They also enable astronomers to detect and study objects that produce little or no light, e.g., black holes, exoplanets or dark matter. For further information, we refer to the monographs [20, 66, 69] and the survey article [65].

In 2012, a galaxy called MACS0647-JD was discovered using the Hubble Space Telescope and the Spitzer Space Telescope[1]. At that time, this was probably the most distant astronomical object we know of. Its light took about 13.3 billion years to reach us. Note that the universe itself is "only" about 13.7 billion years old, so MACS0647-JD belongs to the first generation of galaxies from the very early universe. In this discovery, the galaxy cluster MACS J0647.7+7015 served as a gravitational lens and produced three (magnified) images of MACS0647-JD.

While this thesis was being written, NASA launched the Webb Space

[1] version of February 10, 2022 `https://esahubble.org/news/heic1217/`

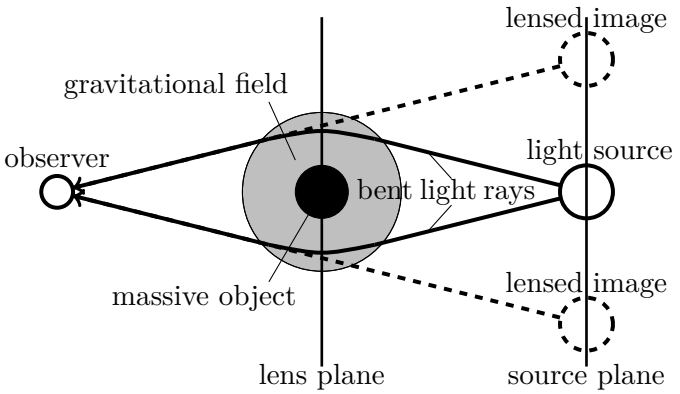


Figure 1: Gravitational lensing with multiple images.

Telescope to observe many more very distant astronomical objects. Gravitational lensing will most likely play an important role for these observations.

The complex formulation of the *n-point mass model* for gravitational lensing leads to the *lens equation*

$$z - \sum_{k=1}^{n} \frac{m_k}{\overline{z} - \overline{z}_k} = \eta. \tag{1.3}$$

The points $z_1, \ldots, z_n \in \mathbb{C}$ are the positions of point-masses $m_1, \ldots, m_n > 0$ in the *lens plane*, which form the gravitational lens, $\eta \in \mathbb{C}$ models the position of a light source in the *source plane*, and the solutions of (1.3) represent the *lensed images* of the light source; see [66, p. 508] or [69, Sect. 8.3.4] for details, and Figure 1 for an illustration with $n = 1$. Obviously, the solutions of (1.3) are the solutions of $f(z) = \eta$ for a harmonic mapping f. Harmonic mappings associated with gravitational lensing are extensively studied in the literature; see, e.g., [9, 11, 12, 26, 37, 41, 47, 70] and the survey articles [10, 39]. It should be noted, however, that in this thesis we also consider harmonic mappings that do not model gravitational lenses.

Of great interest is the behavior of the lensed images under movements of the light source, i.e., changes of the parameter η. This leads to an astrophysical formulation of the recurring questions of this work; see Section 1.1:

Q4: How are the lensed images of a far away light source affected when the light source moves behind a gravitational lens?

Gravitational lensing provides an alternative way to look at the zeros of harmonic mappings. Studies of this effect range from observations with telescopes to theoretical considerations of particular models. We have gained new insights by investigating the following question: What properties of harmonic mappings that model gravitational lensing remain valid for more general harmonic mappings? Of course, observations and special cases can

not replace a rigorous mathematical analysis, but the astrophysics perspective was astonishingly helpful in deriving many of the results of this work, e.g., the possible numbers of zeros for harmonic polynomials (Theorem 3.5), the formula for the number of pre-images (Theorem 4.3), and the overall idea of our continuation method in Section 5.

In the sequel, we will focus on mathematics and let gravitational lensing fade into the background. However, the reader may keep in mind that a far away light source twinkles behind many results in this work.

1.3 Structure of this thesis and main contributions

We organize our original results from [44, 45, 72, 73, 74] into three main topics, each of which is discussed in one of the Sections 3, 4 and 5. Note that this structure is neither "paperwise" nor chronological with respect to the publication date.

Section 2

Section 2 is intended to make the later parts of this work more accessible to the reader. We introduce some basic notation, and briefly review important mathematical tools and results used in this thesis, namely the argument principle for harmonic mappings, results on the dynamics of complex-valued functions, and continuation methods. We would like to emphasize that Section 2 contains no novel results except the complex formulation of Newton's method in Section 2.2 and Theorem 2.4.

Section 3

We present the most important results in the literature regarding bounds on the number of zeros of rational harmonic functions and harmonic polynomials in Section 3.1. In Section 3.2, we summarize our own contributions in [45, 73] on this subject, in particular the sharpened bound on the maximum number of zeros of rational harmonic functions in Theorem 3.2 and Theorem 3.5 on the number of zeros of harmonic polynomials.

Section 4

We briefly review results on the caustics of harmonic mappings in the literature in Section 4.1. After that, in Section 4.2, we present our main results in [44, 72, 73] concerning counting and locating the pre-images of η under a harmonic mapping f: (1) We determine the (global) number of pre-images of η in terms of the winding number of the caustics and the Poincaré indices of the poles. (2) We approximately locate the pre-images of η when $|\eta|$ is sufficiently large. (3) We determine the (local) number of pre-images

of η when η is close to the caustics and also approximately locate these pre-images. Together this gives a complete theory of how the pre-images of η under f depend on the parameter η.

Section 5

In Section 5, we summarize our article [74], in which we developed a continuation method for computing all zeros of (non-singular and non-degenerate) harmonic mappings based on our results in Section 4. Note that the problem of finding *all* zeros is particularly difficult because the number of zeros is usually unknown a priori. We are not aware of any method in the literature that is specialized in this problem for harmonic mappings. Our method works without any prior knowledge about the number or the position of the zeros. We give a complete analysis based on the argument principle and classical convergence results for Newton's method in Section 5.1. Moreover, we provide numerical examples with our `MATLAB` implementation in Section 5.2 in which we observe the following features of our method: It terminates with the correct number of zeros, it is highly accurate in terms of the residual, and it is significantly faster than other (general-purpose) root finders.

Section 6

We conclude with an outlook and a list of possible future research topics that we have encountered in this work.

2 Important mathematical tools used in this work

To show which areas of mathematics have been used in this work and its underlying articles [44, 45, 72, 73, 74], we briefly review the most important results from the literature in this section.

2.1 The argument principle

We start with the argument principle for harmonic mappings, following closely our exposition in [73, Sect. 2.2].

Let f be continuous and non-zero on the trace of a rectifiable curve $\gamma : [a, b] \to \mathbb{C}$. Then the *winding of* f *on* γ is defined as the argument change of $f(z)$ as z travels along γ from $\gamma(a)$ to $\gamma(b)$, divided by 2π, i.e.,

$$W(f;\gamma) = \frac{1}{2\pi}\Delta_\gamma \arg(f(z)) = \frac{1}{2\pi}(\theta(b) - \theta(a)), \tag{2.1}$$

where θ is a continuous branch of the argument of f; see, e.g., [7, Sect. 2.3]. If γ is a closed curve, then the winding is also called the *degree of* f *on* γ and relates to the *winding number* $n(\gamma;\eta)$ *of* γ *about* $\eta \in \mathbb{C} \setminus \operatorname{trace}(\gamma)$ through

$$W(f;\gamma) = n(f \circ \gamma; 0) \quad \text{and} \quad n(\gamma;\eta) = W(z \mapsto z - \eta;\gamma); \tag{2.2}$$

see, e.g., [76, Sect. 2.2.5].

The winding can be visualized using phase plots, in which the domain is colored according to the phase $f(z)/|f(z)|$ of f. We will frequently use this technique, for example in Figure 2, and refer to [83, 84] for a comprehensive discussion of phase plots and their applications. The phase plots in this work were created using the `MATLAB` toolbox of Wegert[2]. To depict also the orientation of f, we use a custom shading. A harmonic mapping f is *sense-preserving/sense-reversing/singular* at z_0, if the *Jacobian of* f,

$$J_f(z) = |\partial_z f(z)|^2 - |\partial_{\overline{z}} f(z)|^2, \tag{2.3}$$

is positive/negative/zero at z_0. Here ∂_z and $\partial_{\overline{z}}$ denote the *Wirtinger derivatives* of f; see, e.g., [24, Sect. 1.2] or [83, p. 144]. Moreover, we call f *singular*, if f is singular at one of its zeros, and *non-singular* (or *regular*) otherwise. In our phase plots, regions where f is sense-preserving, i.e., where the analytic part dominates, are brighter, and regions where f is sense-reversing, i.e., where the anti-analytic part dominates, are darker; see Figure 2. The winding of f on a closed curve γ corresponds to the number of times each color appears in the phase plot while traveling along γ. Therefore it is also called the *chromatic number of* γ; see [83, p. 103]. In the sequel, the winding

[2] version of December 16, 2021, `https://www.mathworks.com/matlabcentral/fileexchange/44375-phase-plots-of-complex-functions`.

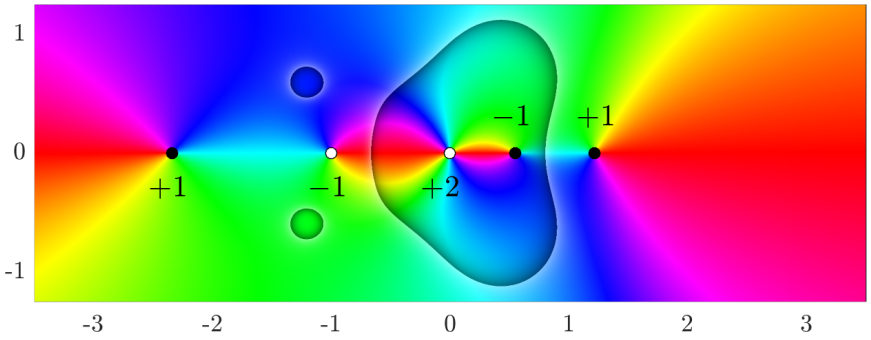


Figure 2: Phase plot of $f(z) = z^2 - 3 + \frac{1}{z+1} - \frac{1}{z} + \overline{\left(z + \frac{1}{z^2}\right)}$ with zeros (black dots) and poles (white dots); Poincaré indices of the exceptional points.

of f on curves where the Jacobian J_f vanishes, will be of particular interest. Three such *critical curves* are apparent in Figure 2.

To determine the winding, we frequently use the following version of Rouché's theorem that is stronger than the classical one for analytic functions. A proof in this more general framework is given in [71, Thm. 2.3].

Theorem 2.1 (Rouché's theorem)**.** *Let γ be a rectifiable closed curve and suppose that f and g are continuous on* $\operatorname{trace}(\gamma)$ *with*

$$|f(z) - g(z)| < |f(z)| + |g(z)| \quad \text{for all } z \in \operatorname{trace}(\gamma). \tag{2.4}$$

Then f and g have the same winding on γ, i.e., $W(f;\gamma) = W(g;\gamma)$.

A point $z_0 \in \mathbb{C}$ is an *isolated exceptional point* of a function f, if f is continuous and non-zero in a punctured disk $D = \{z \in \mathbb{C} : 0 < |z - z_0| < r\}$ for some $r > 0$, and if f is either zero, not continuous, or not defined at z_0. Then the *Poincaré index of f at z_0* is defined as

$$\operatorname{ind}(f; z_0) = W(f;\gamma), \tag{2.5}$$

where γ is an arbitrary positively oriented closed Jordan curve in D surrounding z_0; see [7, p. 39]. The Poincaré index generalizes the multiplicity of zeros and the order of poles of analytic functions; see [76, Sect. 2.5.1]. Similarly, if ∞ is an isolated exceptional point of f, the Poincaré index of f at ∞ can be defined via the transformation $z \mapsto z^{-1}$; see [73, Sect. 2.2].

In a phase plot, the Poincaré index corresponds to the number of times each color appears while traveling once about an exceptional point z_0 in the positive direction. The sign of the index is determined by the order in which the colors appear; see Figure 2.

The argument principle relates the winding of a continuous function to the Poincaré indices of its exceptional points. We frequently used this

concept in [44, 45, 73] to count zeros of harmonic mappings. The following version on the *Riemann sphere* $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ was given by us in [73, Thm. 2.5], and can be obtained from [7, Sect. 2.3] or [76, Sect. 2.3]. Special versions of the argument principle for harmonic mappings are given in [25], [14, Lem. 2.4] and [80, Thm. 2.2].

Theorem 2.2 (Argument principle)**.** *Let D be a multiply connected domain in $\widehat{\mathbb{C}}$ whose boundary consists of Jordan curves $\gamma_1, \ldots, \gamma_n$, which are oriented such that D is on the left. Let f be continuous and non-zero in $\overline{D}$, except for finitely many exceptional points $z_1, \ldots, z_k \in D$, then*

$$\sum_{j=1}^{n} W(f; \gamma_j) = \sum_{j=1}^{k} \operatorname{ind}(f; z_j). \tag{2.6}$$

To illustrate the argument principle, consider again Figure 2. Traveling along the boundary γ of the displayed rectangular domain D in the positive direction, each color appears exactly twice (in the order red, yellow, green), i.e., $W(f; \gamma) = +2$. This corresponds to the sum of the Poincaré indices of the exceptional points of f in D. Since ∞ is the only exceptional point of f in $\widehat{\mathbb{C}}$, that is not in D, we have

$$\operatorname{ind}(f; \infty) = W(f; -\gamma) = -W(f; \gamma) = -2, \tag{2.7}$$

where $-\gamma$ denotes the reversed curve of γ. Note that the Poincaré indices of the exceptional points of f in $\widehat{\mathbb{C}}$ sum up to zero, which also holds in general; see [73, Thm. 2.6].

To apply the argument principle for harmonic mappings, we study the Poincaré indices of their exceptional points, which are either zeros or points at which the function is not defined.

We start with the zeros; see [73, Prop. 2.7], [25, p. 412] and [80, p. 66].

Theorem 2.3. ([73, Prop. 2.7]) *Let f be a harmonic mapping with a zero z_0, such that the local decomposition* (1.2) *is of the form*

$$f(z) = \sum_{k=n}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=n}^{\infty} b_k (z - z_0)^k}, \quad n \geq 1, \tag{2.8}$$

where a_n or b_n can be zero. Then

$$\operatorname{ind}(f; z_0) = \begin{cases} +n, & \text{if } |a_n| > |b_n|, \\ -n, & \text{if } |a_n| < |b_n|, \end{cases} \tag{2.9}$$

and, in particular,

$$\operatorname{ind}(f; z_0) = \begin{cases} +1, & \text{if } f \text{ is sense-preserving at } z_0, \\ -1, & \text{if } f \text{ is sense-reversing at } z_0. \end{cases} \tag{2.10}$$

If $|a_n| = |b_n| \neq 0$, then z_0 is a *singular zero*, i.e., with $J_f(z_0) = 0$. In that case determining the Poincaré index is much more difficult; see, e.g., [46] for harmonic mappings of the form $f(z) = h(z) - \overline{z}$.

Next, we consider points at which f is not defined. If f is harmonic in the punctured disk $D = \{z \in \mathbb{C} : 0 < |z - z_0| < r\}$ and z_0 is an exceptional point of f, then f has a local decomposition

$$f(z) = \sum_{k=-\infty}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=-\infty}^{\infty} b_k (z - z_0)^k} + A \log|z - z_0|, \quad z \in D; \quad (2.11)$$

see [5, Thm. 3], [32, p. 2] or [80, p. 67]. The point z_0 is a *pole* of f if $\lim_{z \to z_0} f(z) = \infty$, and an *essential singularity* if $\lim_{z \to z_0} f(z)$ does not exist in $\widehat{\mathbb{C}}$; see [73, Def. 2.9], [76, p. 44] or [80, Def. 2.1]. For most poles of harmonic mappings, the Poincaré index can be determined from the decomposition (2.11); see [73, Prop. 2.10] and [80, Lem. 2.2, 2.3 and 2.4].

Theorem 2.4. ([73, Prop. 2.10]) *Let f be a harmonic mapping in a punctured neighborhood of z_0, such that the local decomposition* (2.11) *has the form*

$$f(z) = \sum_{k=-n}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=-n}^{\infty} b_k (z - z_0)^k} + A \log|z - z_0|, \quad (2.12)$$

where a_{-n} or b_{-n} can be zero. Then

$$\operatorname{ind}(f; z_0) = \begin{cases} -n, & \text{if } n \geq 1 \text{ and } |a_{-n}| > |b_{-n}|, \\ +n, & \text{if } n \geq 1 \text{ and } |a_{-n}| < |b_{-n}|, \\ 0, & \text{if } n = 0 \text{ and } A \neq 0. \end{cases} \quad (2.13)$$

Note that Theorem 2.4 extends to a possible pole at ∞ using the transformation $z \mapsto z^{-1}$ as described in [73, Sect. 2.2].

In contrast to the argument principle for analytic functions, it is more difficult to obtain the number of zeros of harmonic mappings from the winding, since the Poincaré indices of zeros may cancel out. If the Poincaré indices of all non-singular zeros in a region sum up to $k \in \mathbb{Z}$, we can only conclude that there exists an $n \in \mathbb{N}_0$ such that f has $n + k$ *sense-preserving zeros* (index $+1$) and n *sense-reversing zeros* (index -1). Therefore, we have to count zeros in regions where f has the same orientation at all points. In our publications [44, 73], we have taken this under consideration by applying the argument principle to regions bounded by the already mentioned critical curves.

2.2 Dynamics of complex-valued functions

Next, we review results on the dynamics of complex-valued functions, which we used in our articles [45, 72, 73, 74].

In complex dynamics, we study the iteration of complex-valued functions $f : \Omega \to \Omega$, which are usually analytic; see, e.g., the textbooks [8, 18, 55]

for an introduction. The set Ω is mostly the complex plane $\mathbb{C}$, the Riemann sphere $\widehat{\mathbb{C}}$, or a subset thereof. For $z_0 \in \Omega$, the iteration

$$z_{k+1} = f(z_k), \quad k \geq 0, \tag{2.14}$$

defines a sequence $(z_k)_k$, which is also called the *(forward) orbit of* z_0. Here and in the following, we write $f^1 = f$ and $f^k = f \circ f^{k-1}$ for $k \geq 2$. Of particular interest are *periodic points* of f, i.e., $z_* \in \Omega$ with $f^n(z_*) = z_*$ for some $n \in \mathbb{N}$. For $n = 1$, these are the *fixed points* of f. The *basin of attraction* of a fixed point $z_* \in \Omega$ is defined by

$$A(z_*) = \{z \in \Omega : \lim_{k\to\infty} f^k(z) = z_*\}. \tag{2.15}$$

If $z_0 \in A(z_*)$, i.e., $\lim_{k\to\infty} f^k(z_0) = z_*$, we say that z_* *attracts* z_0 *under* f, or that z_0 *is attracted by* z_* *under* f. Fixed points of analytic functions are classified as follows; see [8, Def. 6.1.1] or [18, p. 27].

Definition 2.5. A fixed point $z_* \in \mathbb{C}$ of an analytic function f is called
(i) *attracting*, if $|f'(z_*)| < 1$,
(ii) *super-attracting*, if $|f'(z_*)| = 0$,
(iii) *repelling*, if $|f'(z_*)| > 1$,
(iv) *rationally neutral*, if $|f'(z_*)| = 1$ and $f'(z_*)^n = 1$ for some $n \in \mathbb{N}$,
(v) *irrationally neutral*, if $|f'(z_*)| = 1$ and $f'(z_*)^n \neq 1$ for all $n \in \mathbb{N}$.

The next theorem on attracting and rationally neutral fixed points of rational functions follows from [18, Ch. III, Thm. 2.2 and 2.3], and may give the reader a basic idea about results in complex dynamics. It is also a main ingredient for the derivation of the sharpened bounds on the maximum number of zeros of rational harmonic functions in Theorem 3.2 below, which is from our article [45].

Theorem 2.6. *Let* $z_* \in \mathbb{C}$ *be an attracting or rationally neutral fixed point of a rational function* r *with* $\deg(r) \geq 2$. *Then there exists a critical point* $z_c \in \mathbb{C}$ *of* r, *i.e.,* $r'(z_c) = 0$, *with* $\lim_{k\to\infty} r^k(z_c) = z_*$.

In other words, every attracting or rationally neutral fixed point of r attracts at least one critical point of r.

Probably the most famous iteration of the form (2.14) is *Newton's method* for analytic functions, which is defined by the iteration

$$z_{k+1} = z_k - \frac{f(z_k)}{f'(z_k)}, \quad k \geq 0; \tag{2.16}$$

see, e.g., [8, Sect. 1.9]. A point $z_* \in \mathbb{C}$ is a simple zero of f, i.e., with $f'(z_*) \neq 0$, if and only if z_* is a (super-attracting) fixed point of the *Newton map*

$$N_f(z) = z - \frac{f(z)}{f'(z)}. \tag{2.17}$$

Newton's method is one of the most widely used iterative root-finding methods and was itself a significant motivation for the development of complex dynamics, starting with the initial works of Schröder and Cayley up to those of Fatou and Julia; see the survey [1] and the monograph [2].

With modern computer graphics, fascinating visualizations of the dynamics of analytic functions became quite popular in the 1980s; see, e.g., the monograph [63]. Images of the *Mandelbrot set*, such as the *Douady rabbit*, gained iconographic status and motivated many people to work on this topic. This led to a *"renaissance of complex dynamics"* [1, Sect. 4].

Also in this work we visualize the dynamics of complex-valued functions and their fixed points, for example in Figure 3. Points which are attracted by the same fixed point have the same color. The darker the shading, the more iterations are required until (numerical) convergence. Points where the iteration does not converge (numerically) are colored in black. This is a commonly used domain coloring technique, especially in the context of Newton's method or Newton-like methods; see, e.g., [28, 82].

Since harmonic mappings are in general not analytic and therefore not differentiable in the complex sense, we cannot use the iteration (2.16) to compute their zeros. However, we can consider Newton's method in a more general framework. Let $F : D \to Y$, $D \subseteq X$ open, be a continuously Fréchet-differentiable map between two Banach spaces X and Y. Then the Newton iteration with initial point $x_0 \in D$ is

$$x_{k+1} = x_k - F'(x_k)^{-1} F(x_k), \quad k \geq 0; \tag{2.18}$$

see, e.g., [23, Sect. 1.2], [33, Ch. 25] or [61, Sect. 7.1]. Under certain regularity conditions on F, the sequence of Newton iterates $(x_k)_k$ is quadratically convergent, i.e., $\|x_{k+1} - x_*\| \leq \text{const} \cdot \|x_k - x_*\|^2$, if x_0 is sufficiently close to a zero x_* of F. The Newton–Kantorovich theorem (see [23, Thm. 2.1] or [61, Thm. 12.6.2]) is one of the classical results where "sufficiently close" is quantified. Here $\overline{B}_r(x_0)$ denotes the closed ball with center x_0 and radius r.

Theorem 2.7 (Newton–Kantorovich)**.** *Let $F : D \to Y$ be a continuously Fréchet-differentiable map with $D \subseteq X$ open and convex. For an initial point $x_0 \in D$ let $F'(x_0)$ be invertible. Suppose that*

$$\|F'(x_0)^{-1} F(x_0)\| \leq \alpha, \tag{2.19}$$

$$\|F'(x_0)^{-1}(F'(y) - F'(x))\| \leq \omega \|y - x\| \quad \text{for all } x, y \in D. \tag{2.20}$$

Let $h = \alpha\omega$ and $\rho = (1 - \sqrt{1 - 2h})/\omega$, and suppose that

$$h \leq \frac{1}{2} \quad \text{and} \quad \overline{B}_\rho(x_0) \subseteq D. \tag{2.21}$$

Then the sequence of Newton iterates $(x_k)_k$ obtained from (2.18) is well-defined, remains in $\overline{B}_\rho(x_0)$, and converges to some x_* with $F(x_*) = 0$. If $h < \frac{1}{2}$, then the convergence is quadratic.

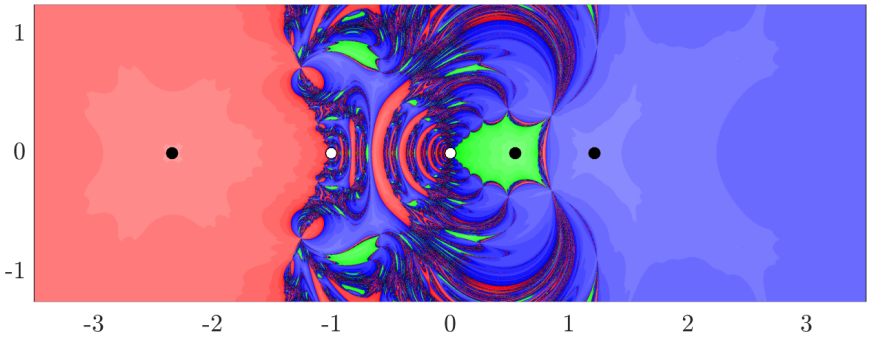


Figure 3: Basins of attraction of the harmonic Newton map H_f for the function $f(z) = z^2 - 3 + \frac{1}{z+1} - \frac{1}{z} + \overline{\left(z + \frac{1}{z^2}\right)}$; zeros (black dots) and poles (white dots) of f; cf. Figure 2.

Harmonic mappings can be regarded as (continuously differentiable) maps from $\mathbb{R}^2$ to $\mathbb{R}^2$. If we reformulate the iteration (2.18) by identifying $\mathbb{R}^2$ with $\mathbb{C}$, and by employing the local decomposition $f = h + \overline{g}$ (see (1.2)), we obtain the *harmonic Newton iteration*

$$z_{k+1} = z_k - \frac{\overline{h'(z_k)} f(z_k) - \overline{g'(z_k) f(z_k)}}{|h'(z_k)|^2 - |g'(z_k)|^2}, \quad k \geq 0; \tag{2.22}$$

see our complete derivation in [72, Sect. 3.1]. The harmonic Newton iteration simplifies to the classical Newton iteration (2.16) when f is analytic ($f = h$), and to $z_{k+1} = z_k - g(z_k)/g'(z_k)$ when f is anti-analytic ($f = \overline{g}$). Thus, it is a natural generalization of Newton's method for analytic functions. The corresponding *harmonic Newton map*

$$H_f(z) = z - \frac{\overline{h'(z)} f(z) - \overline{g'(z) f(z)}}{|h'(z)|^2 - |g'(z)|^2}, \tag{2.23}$$

is in general neither analytic nor harmonic. In particular, the study of its dynamics is not covered by analytic (or anti-analytic) dynamics.

Figure 3 shows the dynamics of H_f for the harmonic mapping f from Figure 2. Black dots mark the (non-singular) zeros of f, which are also fixed points of H_f. The corresponding basins of attraction are colored in red, green and blue. A neighborhood of each zero of f is completely contained in the respective basin of attraction, which is due to the local convergence of Newton's method. In contrast, we see extreme behavior near the poles of f (white dots) and the critical curves (cf. Figure 2).

Indeed, one could argue that the harmonic Newton iteration is nothing but an alternative formulation of Newton's iteration on $\mathbb{R}^2$. But the complex formulation allows to treat harmonic mappings as they are, e.g., without separating real and imaginary parts. As far as we know, the iteration (2.22)

and our generalization for continuously $\mathbb{R}$-differentiable functions (see [83, Def. 4.1.9]),

$$z_{k+1} = z_k - \frac{\overline{\partial_z f(z_k)} f(z_k) - \partial_{\bar{z}} f(z_k) \overline{f(z_k)}}{|\partial_z f(z_k)|^2 - |\partial_{\bar{z}} f(z_k)|^2}, \quad k \geq 0, \tag{2.24}$$

from [72, Sect. 3.1] were unknown in the literature prior to our article [72]. Very recently, our idea was picked up in the preprint [54] to derive the *harmonic Halley method*.

While the local convergence of Newton's method is well understood for fairly general functions going back at least to Kantorovich in the 1940s, the global behavior, except for the special case of analytic functions, is less well understood. In particular, there are only a few results in literature on the global behavior of Newton's method for functions on $\mathbb{R}^2$; see, e.g., [62], [63, Ch. 7], or the recent publications [21, 22]. In addition, there are results on Newton's method on $\mathbb{C}^2$ (e.g. [35]) which of course have implications for the real case. It is worth noting that the iterations (2.22) and (2.24) may lead to an extension of our understanding of the global behavior of Newton's method on $\mathbb{R}^2$ for a broader class of functions, such as harmonic mappings.

2.3 Continuation methods

Finally, we give a brief introduction to the concept of continuation, which we used in [74] to develop a numerical method for computing *all* zeros of certain harmonic mappings. Our presentation is based on [3], [23, Ch. 5] and [61, Sect. 7.5].

Continuation methods are used to solve systems of nonlinear equations. According to [3, p. 1], the idea goes back at least to Poincaré, Klein and Bernstein, and has been applied in a wide range of problems and applications. The typical framework is as follows. Suppose that we have a continuously differentiable map $F : D \to \mathbb{R}^n$, $D \subseteq \mathbb{R}^n$ open, for which the computation of zeros is too "complicated". Instead of solving $F(x) = 0$ directly, we choose an adequate map $G : D \to \mathbb{R}^n$ for which the computation of zeros is "simpler". We then define a *homotopy map*

$$H : D \times [0,1] \to \mathbb{R}^n, \quad \text{with} \quad H(\cdot, 0) = F \quad \text{and} \quad H(\cdot, 1) = G, \tag{2.25}$$

which must also be continuously differentiable. A typical choice is the *convex homotopy*

$$H(x, \lambda) = \lambda G(x) + (1 - \lambda) F(x); \tag{2.26}$$

see [3, Eq. (1.4)] or [61, Sect. 7.5, (2)]. Setting $G(x) = F(x) - d$, $d \in D$, yields the *d-homotopy* [3, Eq. (11.5.2)]

$$H(x, \lambda) = F(x) - \lambda d, \tag{2.27}$$

which is also known as *Newton homotopy*.

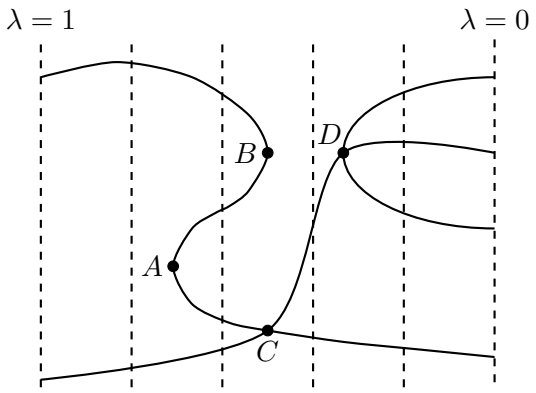


Figure 4: Homotopy curves with turning points A and B, intersection point C, and bifurcation point D.

The idea is now to start at $\lambda = 1$ with the zeros of $G = H(\cdot, 1)$, which are easily accessible by construction. Then we successively compute the zeros of $H(\cdot, \lambda)$ using the previously computed zeros as λ approaches 0. Finally, we obtain the zeros of $F = H(\cdot, 0)$. In other words, we trace the so-called *homotopy curves* $c : [0, 1] \to D$ with $H(c(\lambda), \lambda) = 0$; see Figure 4. Locally, there exists a unique homotopy curve c through a zero (x_k, λ_k) of H if

$$\det \left(\frac{\partial H}{\partial x}(x_k, \lambda_k) \right) \neq 0 \tag{2.28}$$

by the implicit function theorem; see [3, p. 2]. At zeros (x_k, λ_k) of H where the above determinant vanishes, homotopy curves may have turning points, intersection points, or bifurcation points; see Figure 4. Identifying these points and avoiding or properly treating them is one of the biggest challenges in using continuation methods.

One way to trace the implicitly defined homotopy curves (numerically) is a prediction-correction scheme. Such an approach consists of two subroutines (see, e.g., [3, Sect. 2.2] or [23, p. 235]):

1. A *prediction method* that computes from a zero (x_{k-1}, λ_{k-1}) of H a point (s_k, λ_k) that is sufficiently close to the homotopy curve c through (x_{k-1}, λ_{k-1}).

2. A *correction method* that computes from a point (s_k, λ_k) a zero (x_k, λ_k) of H that belongs to the homotopy curve c through (x_{k-1}, λ_{k-1}).

Probably the simplest prediction method is *classical continuation*, where we set $s_k = x_{k-1}$, i.e., we assume that the "old" zero is close to the "new" zero. Another commonly used approach is *tangent continuation*, where we proceed along the tangent to the homotopy curve; see [23, Sect. 5.1.1]. Newton's

method or Newton-like methods are usually used as correction methods; see [23, p. 235] and [3, Ch. 3].

A particular application of the above general framework is the numerical solution of systems of polynomial equations in the field of numerical algebraic geometry. The primary computational methods there are continuation methods, often called *homotopy continuation* in this context; see the monographs [56, 79], the surveys [30, 78], and references therein.

Note that the complex equations, or equivalently, the systems of two real equations, which we solve in Section 5, are in general neither polynomial equations nor can be reformulated into such equations, e.g., for the harmonic mapping (5.8).

3 On the number of zeros of certain harmonic mappings

In this section, we survey bounds on the number of zeros of various classes of harmonic mappings in the literature (Section 3.1) and summarize our own contributions in [45] and [73, Sect. 5] on this subject (Section 3.2).

3.1 Wilmshurst's conjecture and related problems

The fundamental theorem of algebra states that every complex (analytic) polynomial of degree $n \geq 0$ has exactly n zeros in $\mathbb{C}$, counted with their multiplicities. In 1992, Sheil-Small [75] raised the question whether there is an analogue of the fundamental theorem for *harmonic polynomials*, i.e., for functions of the form

$$f(z) = p(z) + \overline{q(z)} = \sum_{j=0}^{n} \alpha_j z^j + \overline{\sum_{j=0}^{m} \beta_j z^j}, \tag{3.1}$$

where p and q are polynomials of respective degrees n and m. This became an active research topic for the following decades with contributions and applications ranging from complex dynamics through algebraic geometry to astrophysics.

Sheil-Small [75] conjectured that every harmonic polynomial as in (3.1) with $n > m \geq 1$, or with $n = m \geq 1$ and $|\alpha_n| \neq |\beta_n|$, has at most n^2 zeros. Recall that "number of zeros" refers to the number of distinct complex points with $f(z) = 0$ in this context, since zeros of harmonic mappings have no multiplicity in the usual sense. Shortly thereafter, Wilmshurst [85, Thm. 5], and independently Peretz and Schmid [64, Cor. 1], confirmed the conjecture of Sheil-Small. Moreover, Wilmshurst proved that the n^2 bound is sharp, by constructing a harmonic polynomial (3.1) with exactly n^2 zeros; see [85, p. 2080] or [76, Sect. 2.6.11]. An alternative proof of the sharpness is due to Bshouty, Hengartner and Suez [15]. In [85, Rem. 2], Wilmshurst conjectured that for $n > m \geq 1$ the number of zeros of every harmonic polynomial is bounded by

$$N(p(z) + \overline{q(z)}) \leq 3n + m(m-1) - 2, \tag{3.2}$$

which coincides with the n^2 bound for $m = n - 1$.

For $m = 1$, Wilmshurst's conjecture (3.2) was settled by Khavinson and Świątek [40, Thm. 1], who proved that harmonic polynomials of the form

$$f(z) = p(z) - \overline{z}, \tag{3.3}$$

with $n = \deg(p) \geq 2$, have at most $3n - 2$ zeros. Their proof relies on the dynamics of polynomial maps and on the argument principle for harmonic mappings (see Section 2.1). However, the question of whether the

bound on the number of zeros of $f(z) = p(z) - \overline{z}$ is sharp remained open. Geyer [27, Thm. 1 and Cor. 2] closed this gap by proving the more general *Crofoot-Sarason conjecture* [16], which states that for all $n \geq 2$ there exist $z_1, \ldots, z_{n-1} \in \mathbb{C}$ and a polynomial p of degree n such that $p(z_j) = \overline{z}_j$ and $p'(z_j) = 0$ hold for $j = 1, \ldots, n-1$. By the argument principle, it follows for all $n \geq 2$ that there exists a polynomial p of degree n such that $f(z) = p(z) - \overline{z}$ has exactly $3n - 2$ zeros.

Replacing the polynomial p in (3.3) by a rational function r, we obtain another type of harmonic mapping for which the maximum number of zeros is extensively studied. These *rational harmonic functions* have the form

$$f(z) = r(z) - \overline{z} \quad \text{with} \quad r(z) = \frac{p(z)}{q(z)}, \tag{3.4}$$

where p and q are coprime polynomials. Using similar techniques as in [40], Khavinson and Neumann [38, Thm. 1] proved that a rational harmonic function (3.4) with $n = \deg(r) \geq 2$ has at most $5n - 5$ zeros, where

$$\deg(r) = \max\{\deg(p), \deg(q)\} \tag{3.5}$$

denotes the *(McMillan) degree* of r. A slightly different proof was given by An and Evans [4, App. C].

Surprisingly, in the context of gravitational lensing (see Section 1.2), the astrophysicist Rhie [67] had already constructed a rational harmonic function (3.4) with $\deg(r) = n$ and exactly $5n - 5$ zeros. However, her point of view was rather different. In the language of astrophysics, she configured an n-point mass lens producing $5n - 5$ lensed images of a single light source. A detailed, rigorous and purely mathematical analysis of Rhie's construction was given by Luce, Sète and Liesen [47, Thm. 2.2]. The same authors sharpened the bound of Khavinson and Neumann by showing in [48, Thm. 1.1] that rational harmonic functions (3.4) with $n = \deg(p) > \deg(q)$ have at most $5n - 6$ zeros. They also proved that this bound is sharp.

While Wilmshurst's conjecture (3.2) is true for $m = 1$ and $m = n - 1$, it does not hold in general. Lee, Lerario and Lundberg [42, Thm. 2.1] gave a counterexample for $m = n-3$. Another counterexample for $m = n-2$ is due to Lee and Saez [43]. Moreover, Hauenstein et al. disproved Wilmshurst's conjecture for further choices of m and n; see [29, Tab. 1]. For this purpose, they computed the zeros numerically by polynomial homotopy continuation and then verified that the computed zeros are actually zeros of the function by using Smale's α-theory [77]. This approach is known as *certified counting*; see [31] and references therein. Khavinson, Lee and Saez [36, Thm. 2 and 3] further contributed to this topic with lower bounds (depending on n and m) on the maximum number of zeros of harmonic polynomials. In a recent preprint, Lundberg [49, Thm. 1.1] used a probabilistic approach to prove that for all $n > m \geq 1$ there exists a harmonic polynomial $f(z) = p(z) + \overline{q(z)}$

with at least $\lceil n\sqrt{m} \rceil$ zeros. This disproves Wilmshurst's conjecture for all $m > 9$ and all sufficiently large n; see [49, Rem. 2].

When Lee, Lerario and Lundberg provided the first counterexample to Wilmshurst's conjecture, they instead conjectured the following bound on the number of zeros

$$N(p(z) + \overline{q(z)}) \leq 2m(n-1) + n; \tag{3.6}$$

see [42, Conj. 1.4] and [49, Sect. 4.1]. But even if this conjecture is true, the above bound would not be sharp for all $m < n$, since the right-hand side in (3.6) exceeds n^2 for $m > \frac{n}{2}$. In any case, the conjectured bound (3.6) is probably not the kind of answer Sheil-Small hoped for when he raised his original question three decades ago. But in general, can we even expect a "Wilmshurst-like" sharp upper bound on the maximum number of zeros of harmonic polynomials? Lee and Saez argue in [43, Rem. 4] why it *"is unlikely based on the known data"* that there exists such a sharp upper bound which is a polynomial in $n = \deg(p)$ and $m = \deg(q)$.

3.2 Our main results

We summarize our own results on the number of zeros of rational harmonic functions and harmonic polynomials from [45] and [73, Sect. 5].

Rational harmonic functions

Let us first collect the bounds on the maximum number of zeros of rational harmonic functions from [38, Thm. 1], [48, Thm. 1.1] and [40, Thm. 1] in the following theorem.

Theorem 3.1. *Let $f(z) = r(z) - \overline{z} = \frac{p(z)}{q(z)} - \overline{z}$ be a rational harmonic function with $n = \deg(r) \geq 2$. Then the number of zeros of f satisfies*

$$N(f) \leq \begin{cases} 5n-5, & \\ 5n-6, & \text{if } \deg(p) > \deg(q), \\ 3n-2, & \text{if } \deg(q) = 0, \text{ i.e., } r \text{ is a polynomial.} \end{cases} \tag{3.7}$$

Although r has degree n in all cases of Theorem 3.1, the (sharp) upper bounds differ significantly. A harmonic polynomial $f(z) = p(z) - \overline{z}$ with $n = \deg(p) \geq 2$ has at most $3n-2$ zeros, while a rational harmonic function of the form

$$f(z) = \frac{p(z)}{\alpha_1 z + \alpha_0} - \overline{z} \tag{3.8}$$

could have up to $5n-6$ zeros according to Theorem 3.1. However, it seems unlikely that the term $\alpha_1 z + \alpha_0$ (instead of 1) in the denominator actually increases the maximum number of zeros by $2n-4$ (from $3n-2$ to $5n-6$).

This motivates us to find better bounds that take the individual degrees of the numerator and denominator polynomials into account.

For this purpose, we consider *shifted* functions $f_\eta = f - \eta$, $\eta \in \mathbb{C}$. Thus, we bound the number of *pre-images* of η under f, which yields also bounds on the number of zeros of f (for $\eta = 0$). On the one hand, considering f_η instead of f fits perfectly with question Q1 in Section 1.1, and on the other hand, it gives better bounds, as we will discuss in item 3 of Remark 3.3.

One of our two main results in [45] is the following theorem. Recall that a harmonic mapping is singular if it has a singular zero and non-singular (or regular) otherwise; see Section 2.1.

Theorem 3.2 ([45, Thm. 1.1])**.** *Let $f(z) = r(z) - \overline{z} = \frac{p(z)}{q(z)} - \overline{z}$ be a rational harmonic function with $\deg(r) \geq 2$. Furthermore, let n_p and n_q be the degrees of p and q, respectively. Then for every $\eta \in \mathbb{C}$, the number of zeros of $f_\eta = f - \eta$ satisfies*

$$N(f_\eta) \leq \begin{cases} 2n_p + 3n_q - 3, & \text{if } n_p < n_q, \\ 5n_p - 5, & \text{if } n_p = n_q, \\ 3n_p + 2n_q - 2, & \text{if } n_p > n_q + 1. \end{cases} \tag{3.9}$$

Our proof in [45] follows the lines of [38] and [4, App. C], and is based on the following observations. Let $r_\eta = r - \eta$, then every zero z_0 of f_η satisfies $r_\eta(z_0) = \overline{z}_0$ and $\overline{r_\eta(z_0)} = \overline{r}_\eta(\overline{z}_0) = z_0$, where $\overline{r}_\eta$ is obtained by complex conjugating all coefficients of r_η. Inserting the first equation into the second yields $\overline{r}_\eta(r_\eta(z_0)) = z_0$. Thus, all zeros of f_η are fixed points of $\overline{r}_\eta \circ r_\eta$ (but not necessarily vice versa), and we can apply Theorem 2.6. Moreover, we use that the set $\{\eta \in \mathbb{C} : f_\eta \text{ is non-singular}\}$ is dense in $\mathbb{C}$; see [38, Fact 1]. Note that the η such that f_η is singular form the *caustics* of f in our later terminology.

Remark 3.3.

1. Our bounds in Theorem 3.2 are better than the best bounds published before (Theorem 3.1) for $n_p < n_q - 1$ and $n_p > n_q + 2$.

2. We omit the case $n_p = n_q + 1$ in Theorem 3.2 since our proof in [45] yields the bound $5n_p - 4$, which is larger than $5n_p - 6$ and therefore not optimal.

3. Separating the constant η from the rational function r in the statement of Theorem 3.2 improves our bounds, as we see next. Let p and q be coprime polynomials with respective degrees n_p and n_q, and with $n_p < n_q - 1$. Furthermore, let $\eta \in \mathbb{C} \setminus \{0\}$ and define

$$r(z) = \frac{p(z)}{q(z)}, \qquad s(z) = \frac{p(z) + \eta q(z)}{q(z)}. \tag{3.10}$$

Then r is of type (n_p, n_q) and s is of type (n_q, n_q). Theorem 3.2 yields the following two different bounds on the number of zeros of the rational harmonic function $r(z) - \overline{z} - \eta = s(z) - \overline{z}$

$$N(r(z) - \overline{z} - \eta) \leq 2n_p + 3n_q - 3, \tag{3.11}$$
$$N(s(z) - \overline{z}) \leq 5n_q - 5. \tag{3.12}$$

The first bound is better than the second, since $n_p < n_q - 1$. Therefore, it is indeed advantageous to separate η from r in Theorem 3.2.

In our second main result in [45], we proved that rational harmonic functions, which are *extremal* in the sense that they attain one of the bounds of Theorem 3.2, are non-singular. This generalizes [48, Thm. 3.1].

Theorem 3.4 ([45, Thm. 3.2])**.** *Let $f(z) = r(z) - \overline{z}$ be a rational harmonic function with $\deg(r) \geq 2$ and $\eta \in \mathbb{C}$. If $f_\eta = f - \eta$ attains one of the bounds of Theorem 3.2, then f_η is non-singular.*

In our proof in [45], we first show that a rational harmonic function f_η with the maximum number of zeros also has the maximum number of *sense-preserving* zeros. Moreover, if f_η is singular, then there exists an $\widetilde{\eta} \in \mathbb{C}$ such that $f_\eta - \widetilde{\eta}$ has more sense-preserving zeros than f_η by [45, Lem. 3.4]. Therefore, extremal rational harmonic functions are non-singular.

The singular zeros of a harmonic mapping f are in a sense the "bad" zeros, since the local behavior of f near these points (e.g. the Poincaré index) is not yet fully understood and still a subject of research; see [46], [52], and also item 4 in the list of possible future research topics in Section 6. Thus, the virtue of Theorem 3.4 is that extremal rational harmonic functions, i.e., with the maximum number of zeros, have only "good" zeros.

Harmonic polynomials

After considering the maximum number of zeros of rational harmonic functions, we return to Sheil-Small's question for a harmonic analogue of the fundamental theorem of algebra from a slightly different perspective. Instead of the maximum number of zeros, we study which numbers of zeros can actually occur for a harmonic polynomial $f(z) = p(z) + \overline{q(z)}$ with $n > m \geq 1$. While every (analytic) polynomial of degree $n \geq 0$ has exactly n zeros (counted with their multiplicities), the situation is more involved for harmonic polynomials. For $n > m \geq 1$ the *maximum valence* of harmonic polynomials

$$V_{n,m} = \max\{N(p(z) + \overline{q(z)}) : \deg(p) = n, \deg(q) = m\} \tag{3.13}$$

is well-defined and finite, since $V_{n,m} \leq n^2$. But it is only known in two special cases, namely $V_{n,n-1} = n^2$ and $V_{n,1} = 3n - 2$; see Section 3.1.

One of our results in [73] concerning the number of zeros of harmonic polynomials is the following.

Theorem 3.5 ([73, Thm. 5.4]). *Let $n > m \geq 1$ and $k \in \{n, n+1, \ldots, V_{n,m}\}$.*

1. *There exists a harmonic polynomial $f(z) = p(z) + \overline{q(z)}$ with $\deg(p) = n$, $\deg(q) = m$, and exactly k zeros.*
2. *If $n - k$ is odd, then every f in item 1 is singular.*
3. *If $n - k$ is even, then there exists a non-singular f in item 1.*

A direct consequence of $V_{n,n-1} = n^2$, $V_{n,1} = 3n - 2$, and item 1 of Theorem 3.5 is the next corollary. It partially generalizes a result of Bleher et al. [12, Thm. 1.1]. Note, however, that [73, Cor. 5.6] contains only item 1 of Corollary 3.6.

Corollary 3.6 ([73, Cor. 5.6]). *Let $n \geq 2$.*

1. *For each $k \in \{n, n+1, \ldots, n^2\}$ there exists a harmonic polynomial $f(z) = p(z) + \overline{q(z)}$ with $\deg(p) = n$, $\deg(q) < n$, and exactly k zeros.*
2. *For each $k \in \{n, n+1, \ldots, 3n-2\}$ there exists a harmonic polynomial $f(z) = p(z) - \overline{z}$ with $\deg(p) = n$ and exactly k zeros.*

Our proofs of the above results in [73] rely mainly on how the number of pre-images of η under f depends on η. We will explore this in great detail in the next section. Surprisingly, harmonic polynomials differing only by an additive constant can be chosen in item 1 of Theorem 3.5 and Corollary 3.6; see [73, Rem. 5.5].

4 Counting and locating pre-images

In this section, we consider the second formulation of our recurring question (Q2: *How do the pre-images of η under f change with η?*). We first briefly recall results on the critical curves and caustics (Section 4.1), and then summarize our main contributions in [44, 72, 73, 74] (Section 4.2). Note that our article [44] is restricted to rational harmonic functions. In [73], we generalized many results from [44] to harmonic mappings. Here we present only the general versions and give references to the special cases in [44].

As we exemplified in Section 1.1, the number of pre-images of η depends highly on η. We analyze this effect and investigate *when*, *where*, and *how* the number of pre-images changes as η varies. We start with an example and consider the harmonic polynomial

$$f(z) = \frac{1}{3}z^3 + \overline{\left(\frac{1}{2}z^2\right)}, \tag{4.1}$$

the rational harmonic mapping

$$f(z) = z - \overline{\left(\frac{z^2}{z^3 - 0.6^3}\right)}, \tag{4.2}$$

and the (transcendental) harmonic mapping

$$f(z) = z^2 + \overline{\left(\frac{1}{z} + \frac{1}{z+1}\right)} + 2\log|z|. \tag{4.3}$$

These functions may look "simple", but it is probably impossible to solve $f(z) = \eta$ for general $\eta \in \mathbb{C}$ by hand. Explicitly determining the number of solutions of $f(z) = \eta$, i.e., the number of pre-images of η, also seems hopeless. For the transcendental function (4.3), not even an upper bound on the *valence*, i.e., the maximum number of pre-images, is known in the literature.

This section is devoted to bring some light into the dark. In particular, we show how the number of pre-images of η under f is related to the position of η with respect to the curves of critical values of f, the so-called *caustics*. This is illustrated in Figure 5, where we see the caustics of the functions (4.1), (4.2) and (4.3). The numbers indicate how many pre-images of η exist when η is in the respective region of the complex plane. We make three important observations:

1. The number of pre-images changes only when η is moved through a caustic of f.

2. The number of pre-images changes exactly by ± 2 when η moves from one side to the other side of a single caustic arc.

3. An η from the region which locally contains the tangent to the caustics has more pre-images.

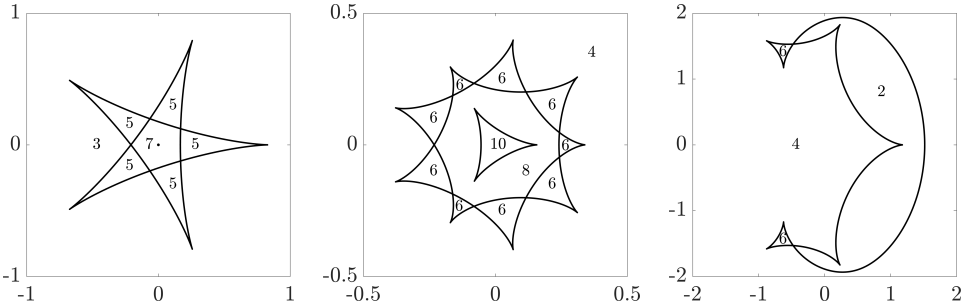


Figure 5: Caustics and number of pre-images of η under the functions (4.1) (left), (4.2) (middle) and (4.3) (right) for an η from the respective regions of the complex plane; cf. [73, Fig. 1, 4 and 6].

Similar observations have been made several times in the literature for certain special classes of harmonic mappings, e.g., for *harmonic trinomials*, $f(z) = z^n + cz^k + 1$, in [13, Fig. 2] and [60, Ex. 5.21], and harmonic mappings of the form $f(z) = z - k/\overline{\sin(z)}, k > 0$, in [11, Fig. 6]. However, a rigorous analysis for general harmonic mappings was lacking before our article [73].

In Section 4.2, we show that our above observations also hold more generally. In particular, in Theorem 4.3 we derive formulas for the number of pre-images of η, which depend crucially on the winding number of the caustics about η. Therefore, we first consider the caustics of harmonic mappings in the next section.

4.1 Critical curves and caustics

In this section, we recall results on the critical curves and the caustics that will be important in the sequel. We closely follow our presentations in [73, Sect. 2.1] and [74, Sect. 2.2], which themselves are based on previous work by Lyzzaik [50, 51, 52] concerning the valence of harmonic mappings.

The *critical set* of a harmonic mapping $f : \Omega \to \mathbb{C}$ consists of the points where the Jacobian of f vanishes, i.e.,

$$\mathcal{C} = \{z \in \Omega : J_f(z) = 0\}. \tag{4.4}$$

In what follows, we assume $J_f \not\equiv 0$, $\partial_{\overline{z}} f \not\equiv 0$, and $\partial_z f \not\equiv 0$. Otherwise, $\mathcal{C} = \Omega$ would hold, f would be analytic, or f would be anti-analytic, respectively. None of this in the scope of this thesis. Under the above assumptions, the set $\mathcal{C}$ consists of isolated points and a level set of the *second complex dilatation of* f

$$\omega(z) = \frac{\overline{\partial_{\overline{z}} f(z)}}{\partial_z f(z)}, \tag{4.5}$$

which is analytic in Ω; see [24, p. 5]. The set of isolated points in $\mathcal{C}$ is

$$\mathcal{M} = \{z \in \Omega : \partial_z f(z) = 0 = \partial_{\overline{z}} f(z) \text{ and } \lim_{\zeta \to z} |\omega(\zeta)| \neq 1\}; \tag{4.6}$$

see [73, Sect. 2.1] and [52, Lem. 2.2]. The remaining part

$$\mathcal{C} \setminus \mathcal{M} = \{z \in \Omega : |\omega(z)| = 1\} \tag{4.7}$$

is a level set of the analytic function ω and therefore consists of analytic curves, which intersect in $z \in \mathcal{C} \setminus \mathcal{M}$ if, and only if $\omega'(z) = 0$. At points $z \in \mathcal{C} \setminus \mathcal{M}$ with $\omega'(z) \neq 0$, the equation

$$\omega(\gamma(t)) = \mathrm{e}^{\mathrm{i}t}, \quad t \in I \subseteq \mathbb{R}, \tag{4.8}$$

implicitly defines a local analytic parametrization $z = \gamma(t)$ of $\mathcal{C} \setminus \mathcal{M}$. Note that f is sense-preserving/sense-reversing to the left/right of γ.

In [73, Sect. 3.1], under the assumption that $\mathcal{C}$ is bounded in $\mathbb{C}$, we proved that $\mathcal{C} \setminus \mathcal{M}$ can be parametrized with analytic closed curves $\gamma_1, \ldots, \gamma_n$ according to (4.8), such that each $z \in \mathcal{C} \setminus \mathcal{M}$ with $\omega'(z) \neq 0$ belongs to exactly one curve γ_j. We call these curves the *critical curves* of f and denote the set of all of them by $\mathrm{crit} = \{\gamma_1, \ldots, \gamma_n\}$; see Figure 6 (top).

Next, we consider the critical values of f, i.e., the image of $\mathcal{C}$ under f. We call $f(\mathcal{C})$ the *set of caustic points*. Let us emphasize the following simple, but important observation: η has a pre-image under f in $\mathcal{C}$ if, and only if η is a caustic point of f. Just like the critical set, $f(\mathcal{C})$ consists of curves and isolated points. For a critical curve $\gamma \in \mathrm{crit}$, we call $f \circ \gamma$ a *caustic* of f. Note, that the caustics are in general not analytic. The caustic points induce a partition of $\mathbb{C} \setminus f(\mathcal{C})$, and we call a connected component $A \subseteq \mathbb{C} \setminus f(\mathcal{C})$ with $\partial A \subseteq f(\mathcal{C})$ a *caustic tile*.

The next lemma characterizes a tangent vector to the caustics and the rate of change of its argument; see [52, Lem. 2.3] for a proof. Its importance cannot be overemphasized, as it is the foundation of many results in this work, e.g., Theorem 4.3 and the continuation method in Section 5.

Lemma 4.1 ([73, Lem. 2.1])**.** *Let f be a harmonic mapping, $z_0 \in \mathcal{C} \setminus \mathcal{M}$ with $\omega'(z_0) \neq 0$, and $z_0 = \gamma(t_0)$ with the parametrization* (4.8)*. Then $f \circ \gamma$ is a parametrization of a caustic and the corresponding tangent vector at $f(z_0)$ is*

$$\tau(t_0) = \frac{\mathrm{d}}{\mathrm{d}t}(f \circ \gamma)(t_0) = \mathrm{e}^{-\mathrm{i}t_0/2}\psi(t_0) \tag{4.9}$$

with

$$\psi(t) = 2\,\mathrm{Re}(\mathrm{e}^{\mathrm{i}t/2}\,\partial_z f(\gamma(t))\,\gamma'(t)). \tag{4.10}$$

In particular, the rate of change of the argument of the tangent vector is

$$\frac{\mathrm{d}}{\mathrm{d}t}\arg(\tau(t))\big|_{t=t_0} = -\frac{1}{2} \tag{4.11}$$

at points with $\psi(t_0) \neq 0$, i.e., the curvature of the caustics is constant with respect to the parametrization $f \circ \gamma$. Moreover, ψ has either only finitely many zeros, or is identically zero, in which case $f \circ \gamma$ is constant.

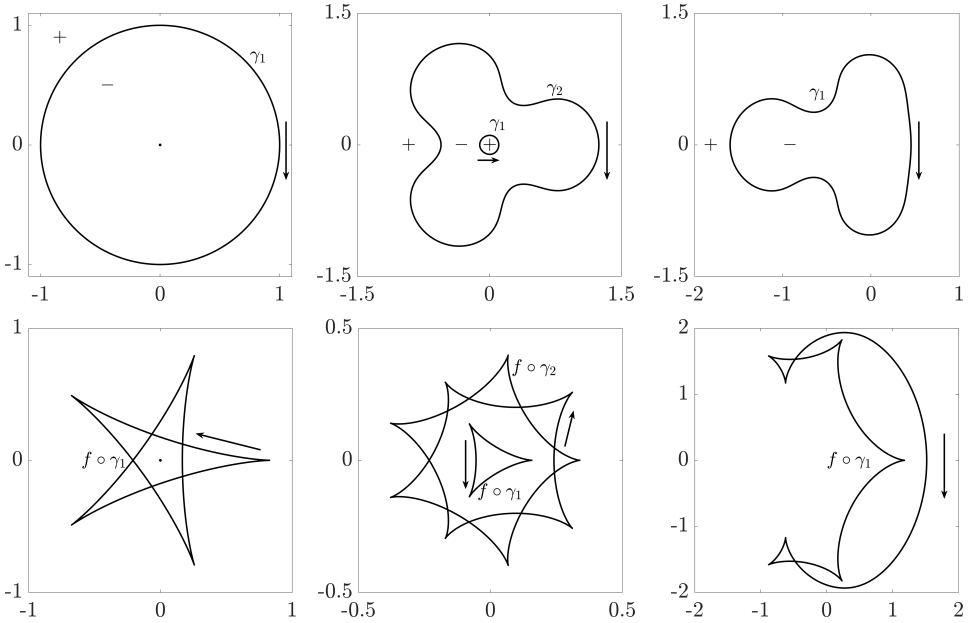


Figure 6: Critical curves (top) and caustics (bottom) for the functions (4.1) (left), (4.2) (middle) and (4.3) (right). Arrows indicate the orientation of the critical curves γ_j and the caustics $f \circ \gamma_j$. The $+/-$ signs indicate the orientation of f in the respective regions. For (4.1), the point $0 \in \mathcal{M}$ is an isolated critical point and $f(0)$ is an isolated caustic point; cf. [73, Fig. 4].

Using the language of *singularity theory* or *catastrophe theory* (see, e.g., Arnold's book [6, Ch. 2]), we call $f(\gamma(t_0))$ a *fold* if the tangent $\tau(t_0)$ is non-zero, i.e., $\psi(t_0) \neq 0$, and a *cusp* if ψ has a zero with sign change at t_0. Folds form arcs, since ψ is continuous. At cusps, the argument of the tangent vector changes by $+\pi$. The caustics and their curvature, caustic tiles, and cusps are apparent in Figures 5 and 6.

Our classification of caustic points is not complete but sufficient for our needs. A similar classification was given by Lyzzaik [52, Def. 2.2]. He has classified the *folding critical points* $z_0 \in \mathcal{C} \setminus \mathcal{M}$ as *critical points* of the *first*, *second* and *third kind*.

4.2 Our main results

We now present our main contributions from [44, 72, 73] concerning counting and locating pre-images under harmonic mappings.

Counting the pre-images

Our first results are restricted to non-degenerate harmonic mappings, a broad class of functions that we introduced in [73, Def. 3.1].

Definition 4.2. We call a harmonic mapping f *non-degenerate* on $\widehat{\mathbb{C}}$ if

1. f is defined in $\widehat{\mathbb{C}}$ except at finitely many poles $z_1, \dots, z_m \in \widehat{\mathbb{C}}$,
2. at each pole $z_j \in \mathbb{C}$ of f, the local decomposition (2.11) of f has the form

$$f(z) = \sum_{k=-n}^{\infty} a_k (z - z_j)^k + \overline{\sum_{k=-n}^{\infty} b_k (z - z_j)^k} + A_j \log|z - z_j| \tag{4.12}$$

   with $n \geq 1$ and $|a_{-n}| \neq |b_{-n}|$, and if $z_j = \infty$ is a pole of f, then

$$f(z) = \sum_{k=-\infty}^{n} a_k z^k + \overline{\sum_{k=-\infty}^{n} b_k z^k} + A_j \log|z|, \quad \text{for } |z| > R, \tag{4.13}$$

   with $n \geq 1$, $|a_n| \neq |b_n|$ and $R > 0$,
3. the critical set of f is bounded in $\mathbb{C}$.

Definition 4.2 guarantees that the argument principle is applicable to f on each connected component $A \subseteq \widehat{\mathbb{C}} \setminus \mathcal{C}$ with $\partial A \subseteq \mathcal{C}$. By item 2, poles are not accumulation points of $\mathcal{C}$. Item 3 implies $J_f \not\equiv 0$, such that $\mathcal{C}$ consists of analytic curves and isolated points. Moreover, the number of critical curves is finite for (non-degenerate) harmonic mappings on $\widehat{\mathbb{C}}$, which is an immediate consequence of [74, Prop. 2.1].

In the context of counting zeros/pre-images, the most intensively studied classes of harmonic mappings are rational harmonic functions $f(z) = r(z) - \overline{z}$ and harmonic polynomials $f(z) = p(z) + \overline{q(z)}$; see Section 3.1. These are non-degenerate if, and only if $\lim_{z\to\infty} |r(z)/z| \neq 1$ and $\lim_{z\to\infty} |p(z)/q(z)| \neq 1$, respectively.

To avoid confusion, we would like to emphasize that the term "non-degenerate" has a slightly different meaning in [44]. There, a rational harmonic function $f(z) = r(z) - \overline{z}$ is non-degenerate if there exists no $z_0 \in \mathcal{C}$ with $r''(z_0) \neq 0$, which implies that critical curves of f do not intersect.

Next, we present our first original result in this section. We give two formulas for the number of pre-images of $\eta \in \mathbb{C} \setminus f(\mathcal{C})$ under a non-degenerate harmonic mapping f, i.e., for

$$N_\eta(f) = |\{z \in \widehat{\mathbb{C}} : f(z) = \eta\}|, \tag{4.14}$$

that confirm our observations at the beginning of this section. Recall that $n(f \circ \gamma; \eta)$ denotes the winding number of the caustic $f \circ \gamma$ about η, and that $\operatorname{ind}(f; z_j)$ denotes the Poincaré index of f at z_j.

Theorem 4.3 ([73, Thm. 3.4 and 3.7])**.** *Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$ with poles $z_1, \dots, z_m \in \widehat{\mathbb{C}}$. Then for all $\eta \in \mathbb{C} \setminus f(\mathcal{C})$ we have*

$$N_\eta(f) = \sum_{j=1}^{m} |\operatorname{ind}(f; z_j)| + 2 \sum_{\gamma \in \text{crit}} n(f \circ \gamma; \eta). \tag{4.15}$$

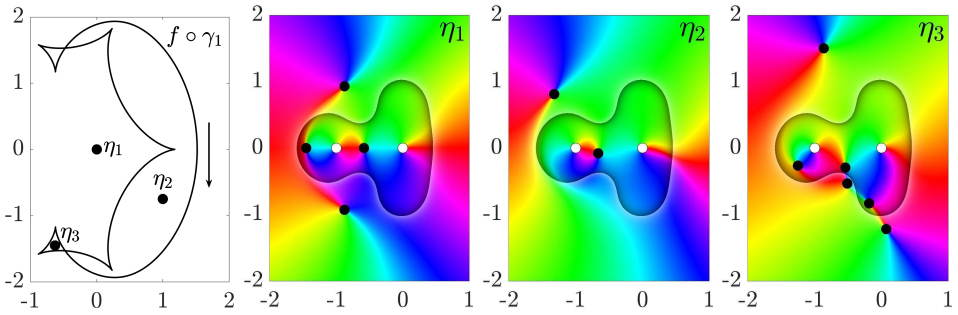


Figure 7: The only caustic $f \circ \gamma_1$ of (4.3) and points η_j. The arrow indicates the orientation of $f \circ \gamma_1$. Phase plots of $f - \eta_j$ with zeros (black dots), i.e., pre-images of η_j under f, and poles (white dots); cf. [74, Fig. 1].

Moreover, if $\eta_1, \eta_2 \in \mathbb{C} \setminus f(\mathcal{C})$ *then*

$$N_{\eta_2}(f) = N_{\eta_1}(f) + 2 \sum_{\gamma \in \text{crit}} \big(n(f \circ \gamma; \eta_2) - n(f \circ \gamma; \eta_1) \big), \tag{4.16}$$

and we have:

1. *If* η_1 *and* η_2 *are in the same caustic tile, then the number of pre-images under* f *is the same, i.e.,* $N_{\eta_2}(f) = N_{\eta_1}(f)$.

2. *If* η_1 *and* η_2 *are separated by a single caustic arc* $f \circ \gamma$*, then the number of pre-images under* f *differs by* 2*, i.e.,* $N_{\eta_2}(f) = N_{\eta_1}(f) \pm 2$.

By a *single* caustic arc, we mean that every $\eta \in \operatorname{trace}(f \circ \gamma)$ has exactly one pre-image in $\mathcal{C}$. In fact, it is possible that distinct critical curves are mapped to the same caustic under f; see our example in [73, Ex. 5.1].

To prove Theorem 4.3, we developed a formula similar to (4.15) for connected components $A \subseteq \widehat{\mathbb{C}} \setminus \mathcal{C}$ with $\partial A \subseteq \mathcal{C}$; see [73, Thm. 3.3]. For this purpose, the argument principle (Theorem 2.2) is applied to $f_\eta = f - \eta$ on A. Since f_η is either sense-preserving or sense-reversing on A, Theorem 2.3 provides the Poincaré indices of its zeros, and since f is non-degenerate, the Poincaré indices of its poles can be determined by Theorem 2.4. Combining the local results we obtain (4.15). The remaining parts of Theorem 4.3 follow immediately from (4.15); see [73, Sect. 3.1] for the details.

Equation (4.16) allows to deduce the number of pre-images of η_2 from the number of pre-images of η_1 if we know the caustics. This partially answers question Q2 in Section 1.1, as it describes how the *number* of pre-images of η changes as η varies. Moreover, equation (4.16) confirms the first and the second observation from the beginning of this section.

To illustrate Theorem 4.3, we again study the non-degenerate and transcendental harmonic mapping (4.3) in Figure 7; see also [73, Ex. 3.10] and [74, Ex. 2.7]. The function f has the poles $z_1 = 0$, $z_2 = -1$ and $z_3 = \infty$ with

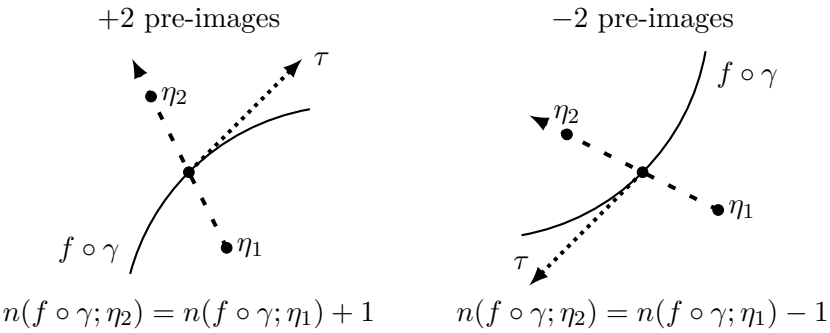


Figure 8: How the curvature of the caustics affects the number of pre-images; see [73, Fig. 3].

Poincaré indices $\mathrm{ind}(f; z_1) = 1$, $\mathrm{ind}(f; z_2) = 1$ and $\mathrm{ind}(f; z_3) = -2$. Since the curvature of a caustic is always negative by Lemma 4.1, the winding numbers of the only caustic $f \circ \gamma_1$ about η_1, η_2 and η_3 are 0, -1 and 1, respectively. According to Theorem 4.3, we have

$$\begin{aligned} N_{\eta_1}(f) &= \sum_{j=1}^{3} |\mathrm{ind}(f; z_j)| + 2n(f \circ \gamma; \eta_1) = 1 + 1 + 2 + 2 \cdot 0 = 4, \\ N_{\eta_2}(f) &= N_{\eta_1}(f) + 2(n(f \circ \gamma; \eta_2) - n(f \circ \gamma; \eta_1)) = 4 + 2(-1 + 0) = 2, \\ N_{\eta_3}(f) &= N_{\eta_1}(f) + 2(n(f \circ \gamma; \eta_3) - n(f \circ \gamma; \eta_1)) = 4 + 2(1 + 0) = 6, \end{aligned}$$

which can be seen in Figure 7.

We can also "geometrically" count the number of pre-images using the caustics, as we elaborated in [73]. By Lemma 4.1, the curvature of a caustic is locally negative. We can therefore read off the orientation of $f \circ \gamma$ from a plot and determine $n(f \circ \gamma; \eta)$ for all $\eta \in \mathbb{C} \setminus f(\mathcal{C})$. In [73, Sect. 3.2], we used the *intersection index* (see also [68, Sect. 3.4]) to confirm the third observation from the beginning of Section 4, namely that an η from the caustic tile that contains locally the tangent to a caustic arc has two more pre-images than an η from the "opposite" side; see Figure 8. We refer to [73, Sect. 3.2] for the details.

Number and location of pre-images for $|\eta|$ sufficiently large

So far we have focused on the *number* of pre-images, but we are also interested in their actual *positions*. We start by considering the pre-images of η when $|\eta|$ is sufficiently large. This is very important for developing the continuation method in Section 5. Our proof in [73] makes extensive use of Rouché's theorem (Theorem 2.1) and the argument principle (Theorem 2.2).

Here and in what follows, we denote by $D_\varepsilon(z_0) = \{z \in \mathbb{C} : |z - z_0| < \varepsilon\}$ the open disk with center $z_0 \in \mathbb{C}$ and radius $\varepsilon > 0$, and by $N_\eta(f; D)$ the number of pre-images of η under f in D.

Theorem 4.4 ([73, Thm. 3.6])**.** *Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$ with poles $z_1, \ldots, z_m \in \widehat{\mathbb{C}}$. Furthermore, let $\varepsilon > 0$ be such that for $j = 1, \ldots, m$ the sets*

$$D_j = \begin{cases} D_\varepsilon(z_j), & \text{if } z_j \in \mathbb{C}, \\ \{z \in \mathbb{C} : |z| > \varepsilon^{-1}\}, & \text{if } z_j = \infty, \end{cases} \tag{4.17}$$

are disjoint, and such that on each set f is either sense-preserving or sense-reversing. Then, for every $\eta \in \mathbb{C}$ with $|\eta|$ sufficiently large, we have

$$N_\eta(f; D_j) = |\mathrm{ind}(f; z_j)| \tag{4.18}$$

for $j = 1, \ldots m$. Moreover, all pre-images of η are in $\bigcup_{j=1}^{m} D_j$.

Note that we slightly reformulated the treatment of a pole at ∞ in Theorem 4.4. However, the statement is the same as in [73, Thm. 3.6].

For rational harmonic functions $f(z) = r(z) - \overline{z}$ we have previously proved an analogous result in [44, Thm. 3.1].

Our next result locates the $|\mathrm{ind}(f; z_j)|$ pre-images of η near a pole $z_j \in \mathbb{C}$ more accurately. Furthermore, it makes the pre-images accessible for numerical computation. Our proof relies on the following strategy: We construct such points that the harmonic Newton iteration (2.22), applied to $f - \eta$ with these initial points, is guaranteed to converge by the Newton–Kantorovich theorem (Theorem 2.7). If all iterates remain in some set, this proves the existence of a pre-image of η under f in that set.

Theorem 4.5 ([74, Thm. 3.2] and [72, Thm. 4.3])**.** *Let f be harmonic in $D = \{z \in \mathbb{C} : 0 < |z - z_0| < r\}$ with a local decomposition (see (2.11))*

$$f(z) = \sum_{k=-n}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=-n}^{\infty} b_k (z - z_0)^k} + A \log|z - z_0|, \quad z \in D, \tag{4.19}$$

where $n \geq 1$ and $|a_{-n}| \neq |b_{-n}|$. Suppose that $c = \eta - (a_0 + \overline{b}_0) \neq 0$, and let $\zeta_1, \ldots, \zeta_n$ be the n solutions of

$$(\zeta - z_0)^n = \frac{|a_{-n}|^2 - |b_{-n}|^2}{\overline{a}_{-n} c - \overline{b}_{-n} \overline{c}}. \tag{4.20}$$

We then have for sufficiently large $|c|$, i.e., for sufficiently large $|\eta|$:

1. *There exist exactly n pre-images of η under f near z_0.*
2. *Each pre-image in item 1 attracts exactly one of the points $\zeta_1, \ldots, \zeta_n$ under the harmonic Newton map $H_{f-\eta}$ from (2.23).*

A similar result holds for a possible pole at ∞; see [72, Cor. 4.5] and also [74, Sect. 3.2] for the details.

Note that the points $\zeta_1, \ldots, \zeta_n$ in Theorem 4.5 are the n solutions of the "truncated" equation $a_{-n}(z - z_0)^{-n} + \overline{b_{-n}(z - z_0)^{-n}} = c$; see [72, Lem. 4.2].

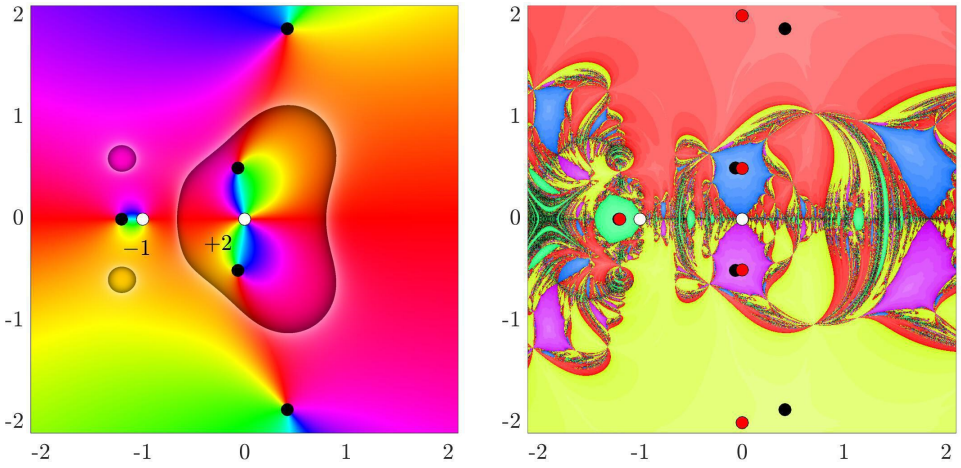


Figure 9: Left: Phase plot of $f-\eta$ for $f(z)=z^2+\frac{1}{z+1}-\frac{1}{z}+\overline{\left(z+\frac{1}{z^2}\right)}$ and $\eta=-3$ with zeros (black dots) and poles (white dots with Poincaré indices). Right: Basins of attraction of $H_{f-\eta}$ with zeros (black dots) and poles (white dots) of $f-\eta$, and the initial points ζ_j in Theorem 4.5 (red dots).

We illustrate Theorems 4.4 and 4.5 in Figure 9. The left plot shows $1=|\mathrm{ind}(f;-1)|$ and $2=|\mathrm{ind}(f;0)|$ pre-images of η under f near the poles -1 and 0, respectively. The two remaining pre-images are "close" to the pole at ∞ with $|\mathrm{ind}(f;\infty)|=2$. This confirms Theorem 4.4. The right plot shows the basins of attraction of the harmonic Newton map $H_{f-\eta}$. Each pre-image of η (black dots) attracts exactly one initial point ζ_j (red dots), as stated in Theorem 4.5. Note that the pre-image and the corresponding initial point near the pole -1 are almost identically, so that the black dot is barely visible.

Crossing a caustic arc at a fold

Next, we investigate the (local) effect on the pre-images of η, when η "moves" through a caustic arc. We call this a *caustic crossing*. By Theorem 4.3, the number of pre-images of η then changes by ± 2. We study whether we have additional or fewer pre-images. In the case of an increasing number of pre-images, we locate "new" pre-images and make them accessible for numerical computation. For simplicity, we only consider caustic crossings through folds in a certain direction c. This is sufficient to develop our continuation method in Section 5.

Theorem 4.6 ([73, Thm. 4.2])**.** *Let f be a harmonic mapping and $z_0 \in \mathcal{C}\setminus\mathcal{M}$, such that $\eta=f(z_0)$ is a fold caustic point. Moreover, let*

$$f(z)=\sum_{k=0}^{\infty}a_k(z-z_0)^k+\overline{\sum_{k=0}^{\infty}b_k(z-z_0)^k} \quad \textit{and} \quad c=-\left(\frac{a_2\bar{b}_1}{a_1}+\frac{\bar{b}_2 a_1}{\bar{b}_1}\right). \tag{4.21}$$

Then, for each sufficiently small $\varepsilon > 0$, there exists a $\delta > 0$, such that for all $0 < t < \delta$:

1. *$\eta - tc$ has no pre-image under f in $D_\varepsilon(z_0)$,*
2. *η has exactly one pre-image under f in $D_\varepsilon(z_0)$,*
3. *$\eta + tc$ has exactly two pre-images under f in $D_\varepsilon(z_0)$.*

In item 3, f is sense-preserving at one pre-image and sense-reversing at the other. Moreover, if $t > 0$ is sufficiently small, then each pre-image in item 3 attracts one of the points

$$\zeta_\pm = z_0 \pm \mathrm{i}\sqrt{t\frac{\overline{b}_1}{a_1}} \tag{4.22}$$

under the harmonic Newton map $H_{f-(\eta+tc)}$.

Since Theorem 4.6 describes a local effect, it is not necessary that f is non-degenerate or even globally defined. Note that in [73, Thm. 4.2] we assumed, as did Lyzzaik in [52, p. 135], that f is (locally) *light*, i.e., the pre-image $f^{-1}(\{\eta\})$ is totally disconnected for all η. However, since a harmonic mapping f is always light in a neighborhood of $z_0 \in \mathcal{C}$ if $f(z_0)$ is a fold, we skip this assumption here; see [73, Rem. 2.3].

Our proof in [73] has three main ingredients. (1) By our previous result [72, Thm. 5.2 (1)], which is based on the Newton–Kantorovich theorem (Theorem 2.7), the pre-images in item 3 attract the points $\zeta_\pm$ under $H_{f-(\eta+tc)}$. In particular, this proves the existence of the two pre-images of $\eta + tc$ in $D_\varepsilon(z_0)$. Note that similar to the proof of Theorem 4.5, the points $\zeta_\pm$ are the solutions of the "truncated" equation

$$\sum_{k=0}^{2} a_k(z - z_0)^k + \overline{\sum_{k=0}^{2} b_k(z - z_0)^k} = \eta + tc; \tag{4.23}$$

see [72, Sect. 5] for more details. (2) By Lyzzaik's study on the valence of light harmonic mappings [52, Thm. 5.1], all $\widetilde{\eta}$ have at most two pre-images in $D_\varepsilon(z_0)$. (3) Finally, by Theorem 4.3 for non-degenerate harmonic mappings, or by Neumann's work for more general functions [60, Thm. 6.7], $\eta - tc$ has no pre-image in $D_\varepsilon(z_0)$.

We gave a similar result for rational harmonic functions $f(z) = r(z) - \overline{z}$ in [44, Thm. 4.3], but in contrast to [73, Thm. 4.2], its proof relies mainly on Rouché's theorem (Theorem 2.1) and the argument principle (Theorem 2.2).

While the statement of Theorem 4.6 is restricted to the direction c in (4.21), we have a similar effect on the number of pre-images in $D_\varepsilon(z_0)$ for all other directions that are not tangential to the caustics; see our discussion in [73, Rem. 4.3].

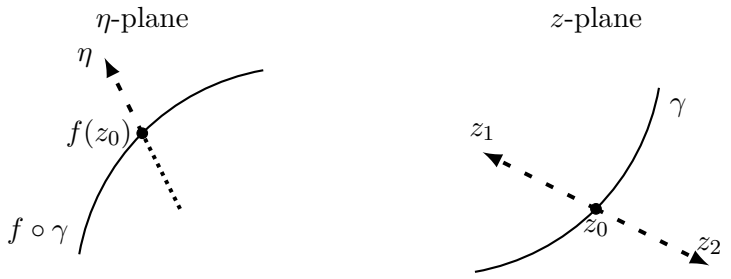


Figure 10: Local behavior when η crosses a fold; cf. [73, Fig. 5].

If a caustic arc is crossed at a cusp instead of a fold, we proved a somewhat weaker result in [72, Thm. 5.2 (2)]; see also [44, Thm. 4.5] for rational harmonic functions.

Theorem 4.6 is illustrated in Figures 10 and 11. Figure 10 shows a schematic illustration of a caustic crossing at a fold. While η moves along the dotted line, no pre-image of η exists near z_0. When η reaches the caustic at $f(z_0)$, a (singular) pre-image of η appears at the corresponding critical point z_0, which then branches into two (non-singular) pre-images z_1 and z_2 as η moves away from the caustic along the dashed line. Figure 11 shows how the phase plots and the basins of attraction change during a caustic crossing. Interestingly, the future basins of attraction are already visible in the bottom left plot as black regions, i.e., where the harmonic Newton iteration (2.22) does not numerically converge within a given number of steps. (For these plots, the maximum number of iterations was set to 25.) In the bottom middle plot, we see the basin of attraction of the singular pre-image 0 of η (cyan), which is shaded very dark. This indicates a fairly slow convergence of the iteration (2.22), as we might expect for a singular pre-image. Recall, the darker the shading, the more iterations are required. When this singular pre-image branches into (non-singular) pre-images of $\eta + tc$, the convergence speed increases.

In the Theorems 4.4, 4.5 and 4.6, the terms "sufficiently small/large" may sound vague and unsatisfactory, especially if we are interested in practical computations. However, as we will see in the next section, we can use our results to develop an extremely fast and highly accurate numerical method for computing *all* zeros of a non-singular and non-degenerate harmonic mapping.

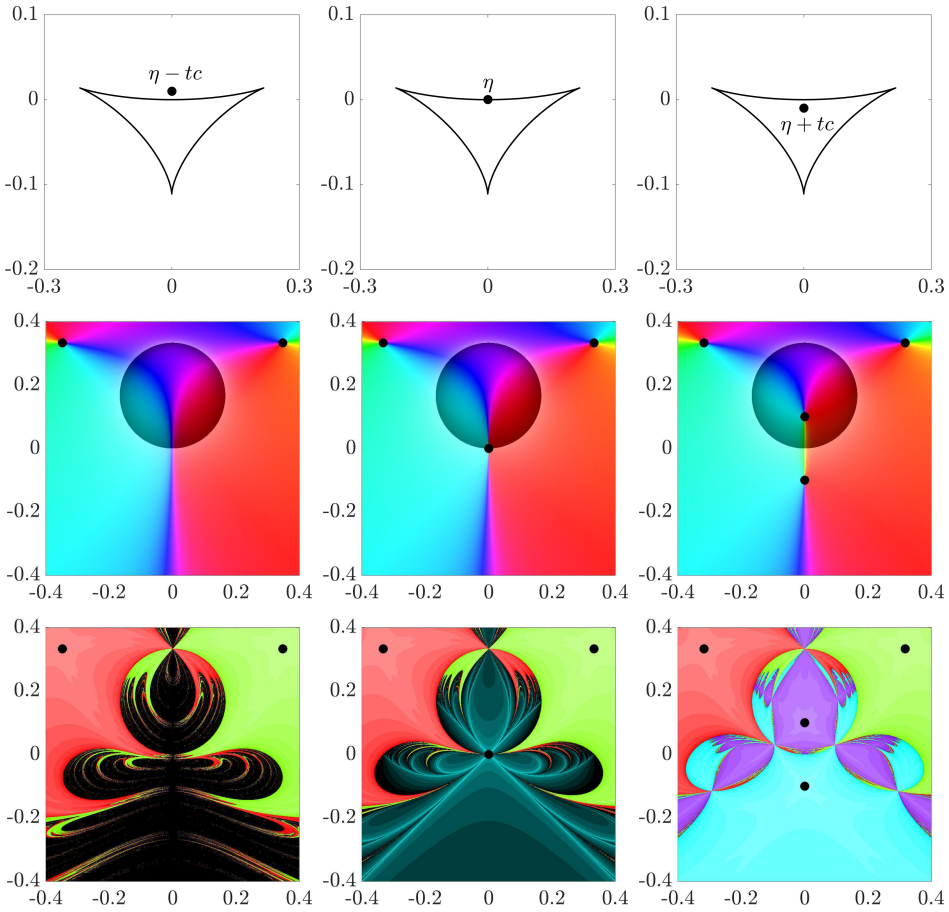


Figure 11: Top: The only caustic of $f(z) = 2\mathrm{i}z^2 + z + \overline{\mathrm{i}z^2 + z}$ and the points $\eta - tc$, η and $\eta + tc$ for $\eta = 0$, $c = -\mathrm{i}$, $t = 0.01$ and $z_0 = 0$ as in Theorem 4.6. Middle: Phase plots of $f - (\eta - tc)$, $f - \eta$ and $f - (\eta + tc)$ with zeros (black dots). Bottom: Basins of attraction of $H_{f-(\eta-tc)}$, $H_{f-\eta}$ and $H_{f-(\eta+tc)}$.

5 Numerical computation of zeros by continuation

After our theoretical results, we now turn to the numerical computation of zeros of harmonic mappings and summarize our main contributions in [74]. In doing so, we consider the third formulation of our recurring question (Q3: *How do the solutions of $f(z) = \eta$ depend on η?*) and closely follow our presentation in [74, Sect. 3].

A *single* zero of a harmonic mapping can be computed by any adequate iterative root finder, e.g., our formulation of Newton's method in Section 2.2. The computation of *all* zeros is much more challenging and we are not aware of any method for this problem in the literature that is specialized to harmonic mappings. Of course, one idea is to run a root finder with multiple starting points. However, it is by no means trivial to find a set of starting points which guarantees that all zeros are computed. In their groundbreaking article [34], Hubbard, Schleicher and Sutherland constructed a set $\mathcal{S}_n$ of approximately $1.11n\log(n)^2$ points, such that each zero of any (suitably normalized) polynomial p of degree n attracts an $s \in \mathcal{S}_n$ under the respective Newton map N_p (see (2.17)). For harmonic mappings this problem is even harder, especially since the number of zeros is in general unknown a priori, as we have elaborated in the previous sections.

In Section 5.1 we use a different approach which is based on continuation. We develop and analyze the *transport of images method* for computing *all* zeros of a non-degenerate and non-singular harmonic mapping. Moreover, in Section 5.2, we present numerical examples with our `MATLAB` implementation from [74].

5.1 The transport of images method

The name of our method is borrowed from the theory of gravitational lensing in astrophysics; see Section 1.2. In this context, Schneider, Ehlers and Falco [69, Sect. 10.5] have sketched a continuation approach to solve the lens equation. However, almost no details about analysis and implementation are given. In fact, they themselves are rather pessimistic about an actual implementation of the idea: *"But there are no guaranteed methods to find all the roots in a given part of the lens plane."* [69, p. 300]. Although the name "transport of *pre*-images" would be mathematically more precise, we kept the naming of Schneider, Ehlers and Falco. For them, pre-images of η are lensed images of a light source at a position modeled by η.

After removing the astrophysical overhead, the basic idea of the method is rather simple. For a harmonic mapping f, we first solve $f(z) = \eta_1$ for some $\eta_1 \in \mathbb{C}$. We then construct a sequence $\eta_2, \eta_3, \ldots$ and solve $f(z) = \eta_{k+1}$ based on the solutions of $f(z) = \eta_k$. We let the sequence $(\eta_k)_k$ tend towards the origin and eventually, after finitely many steps, we end up with $\eta_n = 0$, so that solving $f(z) = \eta_n$ yields all zeros of f. Note that "solve" here

means the numerical computation of *all* solutions. Embedded in the theory of continuation methods (see Section 2.3 for a brief introduction), this is nothing else than continuation with d-homotopy (or Newton homotopy).

Our main contribution in [74] is that we complement this basic idea with a rigorous mathematical analysis and a precise algorithmic description. For this purpose, we use a prediction-correction scheme with an adapted version of classical continuation as prediction and the harmonic Newton method from Section 2.2 as correction. Our method works without any prior knowledge about the number or the position of the zeros. Moreover, we prove that *all* zeros are computed with finitely many continuation steps.

The main novelty and the key feature is that we overcome two typical difficulties encountered in continuation methods:

1. *Computing all initial solutions.*

   First of all, it is unclear why it should be simpler to compute all solutions of $f(z) = \eta_1$ for any $\eta_1 \in \mathbb{C}$ than it would be to compute all zeros of f. But using our theoretical results in Section 4.2, we can easily compute all solutions of $f(z) = \eta$ when $|\eta|$ is sufficiently large.

2. *Determining when, where and how the number of solutions changes.*

   Recall that the number of solutions of $f(z) = \eta_k$ and $f(z) = \eta_{k+1}$ may differ if the homotopy curves (see Figure 4) have turning, bifurcation or intersection points. Based on our theoretical results in Section 4.2, we can properly treat turning points and, moreover, avoid the (finitely many) bifurcation and intersection points.

In both cases, we construct a (minimal) set of points S_k such that applying the harmonic Newton iteration (2.22) on S_k yields *all* solutions of $f(z) = \eta_k$. This motivates the next definition; see [74, Def. 3.1].

Definition 5.1. Let f be a harmonic mapping and $\eta \in \mathbb{C} \setminus f(\mathcal{C})$. For all initial points z_0 in a neighborhood of a (non-singular) solution z_* of $f(z) = \eta$, the limit of the harmonic Newton iteration for $f - \eta$ exists by the local convergence property of Newton's method, and we can define

$$H^{\infty}_{f-\eta}(z_0) = \lim_{k\to\infty} H^{k}_{f-\eta}(z_0). \tag{5.1}$$

We call S a *prediction set* of the pre-image $f^{-1}(\{\eta\})$ if

$$H^{\infty}_{f-\eta}(S) = f^{-1}(\{\eta\}) \quad \text{and} \quad |S| = |f^{-1}(\{\eta\})|. \tag{5.2}$$

If S is a prediction set, then the map $H^{\infty}_{f-\eta} : S \to f^{-1}(\{\eta\})$ is bijective, i.e., each point $z_0 \in S$ can be identified with a solution of $f(z) = \eta$, and vice versa. We have already seen an example of a prediction set in Figure 9 (right plot). Smale [77, p. 186] calls a point $z_0 \in S$ an *approximate zero* of $f - \eta$, if the Newton iterates additionally satisfy $|z_k - z_{k-1}| \leq (\frac{1}{2})^{2^{k-1}-1}|z_1 - z_0|$.

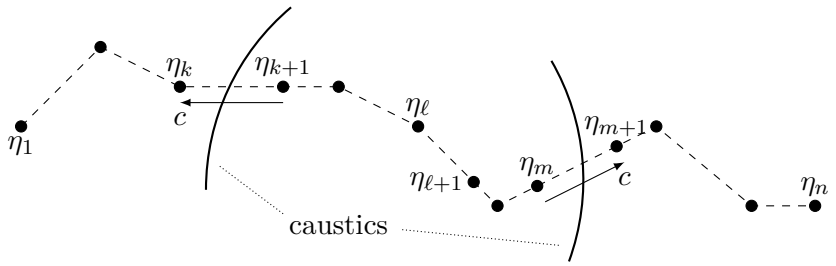


Figure 12: A path $(\eta_1,\ldots,\eta_n) \in \mathbb{C}^n$ with the three kinds of steps in the transport of images method: (1) $f(z) = \eta_{k+1}$ has two solutions fewer than $f(z) = \eta_k$, (2) the numbers of solutions of $f(z) = \eta_\ell$ and $f(z) = \eta_{\ell+1}$ are the same, and (3) $f(z) = \eta_{m+1}$ has two solutions more than $f(z) = \eta_m$. The arrows indicate the direction c in Theorem 4.6; see [74, Fig. 2].

Using the notion of a prediction set, we obtain the following description of our method.

The transport of images method

1. *Initial phase:* Construct a prediction set S_1 of $f^{-1}(\{\eta_1\})$ for some η_1 and apply $H_{f-\eta_1}$ on S_1 to compute all solutions of $f(z) = \eta_1$.
2. *Transport phase:* While $\eta_k \neq 0$ do
   (a) *Prediction:* Choose an η_{k+1} and construct a prediction set S_{k+1} of $f^{-1}(\{\eta_{k+1}\})$ from the solutions of $f(z) = \eta_k$.
   (b) *Correction:* Apply $H_{f-\eta_{k+1}}$ on S_{k+1} to compute all solutions of $f(z) = \eta_{k+1}$.

Next, we describe how prediction sets can be constructed in the initial phase and in the transport phase.

Initial phase

For $|\eta_1|$ sufficiently large (in particular, such that η_1 is outside of all caustics), Theorem 4.4 states that all solutions of $f(z) = \eta_1$ lie near to the finitely many poles of f (which we assume to be known). By Theorem 4.5, the set of all ζ_j (solutions of (4.20)) for all poles of f is a prediction set of $f^{-1}(\{\eta_1\})$; see [74, Sect. 3.2] for more details.

Transport phase

We distinguish the steps from η_k to η_{k+1} into those where η_k and η_{k+1} are in the same caustic tile, and those where η_k and η_{k+1} are separated by a caustic arc; see Figure 12.

If η_k and η_{k+1} are in the same caustic tile and the stepsize $|\eta_{k+1} - \eta_k|$ is sufficiently small, we can use classical continuation, i.e., the "old" solutions are a prediction set of the "new" solutions.

Theorem 5.2 ([74, Thm. 3.4])**.** *Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$ and let $\eta_k \in \mathbb{C} \setminus f(\mathcal{C})$. Then there exists an $\varepsilon > 0$, such that*

$$S = f^{-1}(\{\eta_k\}) \tag{5.3}$$

is a prediction set of $f^{-1}(\{\eta_{k+1}\})$ for all $\eta_{k+1} \in D_\varepsilon(\eta_k) \subseteq \mathbb{C} \setminus f(\mathcal{C})$.

Our proof of the above result in [74] is based on the Newton–Kantorovich theorem (Theorem 2.7) and the inverse function theorem.

If η_k and η_{k+1} are separated by a single caustic arc, the number of solutions of $f(z) = \eta_k$ and $f(z) = \eta_{k+1}$ differs by ± 2. Therefore, we need to add or remove two points from $f^{-1}(\{\eta_k\})$ to obtain a prediction set of $f^{-1}(\{\eta_{k+1}\})$. Due to the next theorem, this can be achieved if the step from η_k to η_{k+1} is done in a certain direction. Here and in what follows, we call a caustic point η *simple* if $|f^{-1}(\{\eta\}) \cap \mathcal{C}| = 1$.

Theorem 5.3 ([74, Thm. 3.6])**.** *Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$ and $\eta_k, \eta_{k+1} \in \mathbb{C} \setminus f(\mathcal{C})$, such that there exists exactly one simple fold η and no other caustic point on the line segment from η_k to η_{k+1}. Moreover, let c and $\zeta_\pm$ be defined as in Theorem 4.6 (see* (4.21) *and* (4.22)*), and suppose that* $\arg(\eta_{k+1} - \eta_k) = \arg(\pm c)$ *holds. If the stepsize $|\eta_{k+1} - \eta_k|$ is sufficiently small, then*

$$S = \begin{cases} f^{-1}(\{\eta_k\}) \cup \{\zeta_\pm\}, & \text{if } \arg(\eta_{k+1} - \eta_k) = \arg(+c), \\ f^{-1}(\{\eta_k\}) \setminus H^\infty_{f-\eta_k}(\{\zeta_\pm\}), & \text{if } \arg(\eta_{k+1} - \eta_k) = \arg(-c), \end{cases} \tag{5.4}$$

is a prediction set of $f^{-1}(\{\eta_{k+1}\})$.

Our proof in [74] crucially relies on our study of caustic crossings at folds in Theorem 4.6 and again on the Newton–Kantorovich theorem.

We illustrate Theorems 5.2 and 5.3 in Figure 13. For the step from η_1 to η_2, each basin of attraction of $H_{f-\eta_2}$ contains exactly one point of $f^{-1}(\{\eta_1\})$, i.e., $f^{-1}(\{\eta_1\})$ is a prediction set of $f^{-1}(\{\eta_2\})$. The step from η_2 to η_3 crosses a caustic at a simple fold. The curvature implies that $f(z) = \eta_3$ has two more solutions than $f(z) = \eta_2$. The set $S = f^{-1}(\{\eta_2\}) \cup \{\zeta_\pm\}$ from Theorem 5.3 is a prediction set of $f^{-1}(\{\eta_3\})$. Note that the points $\zeta_\pm$ are barely visible because they almost coincide with the corresponding "new" solutions of $f(z) = \eta_3$.

Transport paths

Along a sequence $(\eta_k)_k$ where Theorem 5.2 or Theorem 5.3 holds for all steps from η_k to η_{k+1}, we can "transport" the solutions. This motivates our next definition; see [74, Def. 3.9].

Definition 5.4. Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$. We call $(\eta_1, \ldots, \eta_n) \in \mathbb{C}^n$ a *transport path* from η_1 to η_n, if for all $k = 1, \ldots, n-1$ either S as in (5.3) or S as in (5.4) is a prediction set of $f^{-1}(\{\eta_{k+1}\})$.

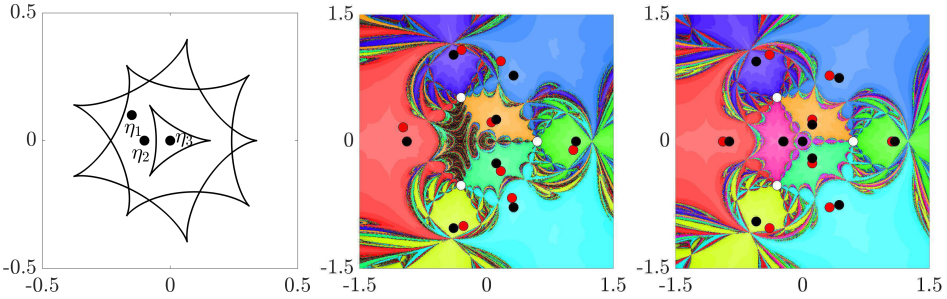


Figure 13: Illustration of Theorems 5.2 and 5.3 for $f(z) = z - \overline{z^2/(z^3 - 0.6^3)}$. Left: Caustics and η_j. Middle: Basins of attraction of $H_{f-\eta_2}$, prediction set $f^{-1}(\{\eta_1\})$ (red dots) of $f^{-1}(\{\eta_2\})$ (black dots), and poles of f (white dots). Right: Basins of attraction of $H_{f-\eta_3}$, prediction set $f^{-1}(\{\eta_2\}) \cup \{\zeta_\pm\}$ (red dots) of $f^{-1}(\{\eta_3\})$ (black dots), and poles of f (white dots); see [74, Fig. 4].

In [74, Thm. 3.10], we proved the existence of transport paths under very mild assumptions. We call η a *multiple* caustic point, if $|f^{-1}(\{\eta\}) \cap \mathcal{C}| \geq 2$. As mentioned earlier, it is indeed possible that a harmonic mapping f maps two critical curves to the same caustic; see our example in [73, Ex. 5.1].

Theorem 5.5 ([74, Thm. 3.10])**.** *Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$ without non-isolated multiple caustic points and η_s, $\eta_e \in \mathbb{C} \setminus f(\mathcal{C})$. Then there exists a transport path from η_s to η_e.*

Our proof in [74] consists of four steps. (1) We prove that f has only finitely many caustic points that are not simple folds. (2) Next, we show that there exists a (rectifiable) polygonal path from η_s to η_e with finitely many caustic crossings, all at simple folds. (3) With arbitrarily small manipulations of this path, we ensure that each caustic crossing is done in the direction $\pm c$ in (4.21). (4) Finally, we refine the path by introducing finitely many intermediate points to obtain a transport path.

The next corollary follows from Theorem 5.5 and our description of the initial phase. Recall that f is singular, if and only if 0 is a caustic point.

Corollary 5.6 ([74, Cor. 3.12])**.** *Let f be a non-singular and non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$ without non-isolated multiple caustic points. Then there exists an $\eta_1 \in \mathbb{C} \setminus f(\mathcal{C})$, such that Theorem 4.5 applies to all poles of f, and such that there exists a transport path from η_1 to 0.*

Note that we excluded singular zeros in our theoretical results and also in the numerical examples that follow. Since continuation usually encounters numerical problems when solutions are singular or nearly singular, special strategies are used, often referred to as the *endgame*. Two classical strategies for systems of analytic functions are the *Cauchy endgame* [57] and the *power series endgame* [58]. We very briefly discussed endgame issues for the transport of images method in [74, Rem. 3.13]; see also [74, Sect. 3.4].

5.2 Implementation and numerical examples

In this section, we present some of our numerical examples in [74, Sect. 5]. For this purpose, we use our `MATLAB` implementation of the transport of images method, which can be downloaded from `GitHub`:

```
https://github.com/transportofimages/Transport_of_images_Toolbox
```

It contains routines for computing the zeros and caustics of non-singular and non-degenerate harmonic mappings. Details on the implementation can be found in [74, Sect. 4] and in the `MATLAB` code itself. However, one particular aspect of our implementation is so crucial that we would like to discuss it briefly here as well; see also [74, Sect. 4.2].

Construction of transport paths

In the transport phase, it is theoretically essential to cross the caustics in the direction $\pm c$ (see (4.21)), but empirically this seems to be less important in many examples. Therefore, we do not consider this assumption in our implementation. However, if for each step from η_k to η_{k+1} the set S constructed as in (5.3) or (5.4), but possibly with $\arg(\eta_{k+1} - \eta_k) \neq \arg(\pm c)$, is a prediction set of $f^{-1}(\{\eta_{k+1}\})$, we still call $(\eta_1, \dots, \eta_n) \in \mathbb{C}^n$ a transport path; cf. Definition 5.4.

For simplicity, we construct transport paths in our implementation on rays $R_\theta = \{r\mathrm{e}^{\mathrm{i}\theta} : r \geq 0\}$, $\theta \in \mathbb{R}$. Then the intersection points of R_θ and the caustics can be read off from the argument of the caustics. However, it may be difficult to determine an angle θ such that there exists a transport path on R_θ. Problems arise especially when R_θ is tangential or almost tangential to the caustics, or when R_θ intersects the caustic at or near non-fold points. To overcome this difficulty, we draw $\theta \in]0, 2\pi[$ uniformly distributed with the `rand` routine in `MATLAB`. If the transport phase with θ is not successful we restart with a new random angle.

While continuation methods typically use *adaptive step size control* (see, e.g., [3, Ch. 6]), we use here a different approach that is best described as *greedy divide-and-conquer*. The term "greedy" means that we try to make the steps as large as possible, and "divide-and-conquer" means that we divide a step into substeps when necessary. We illustrate the idea in Figure 14; see also our formal description in [74, Sect. 4.2]. For each intersection point of R_θ and the caustics, we choose two points η_{2k} and η_{2k+1} (one of each side of the respective caustic arc). These points, together with η_1 from the initial phase and $\eta_n = 0$, form the *points of depth* 0 of our (potential) transport path. Then we try to "transport" the solutions. Whenever a transport step fails, we divide it into substeps by inserting *points of depth* 1 to our path. We proceed recursively until we find a transport path or until we reach a prescribed recursion depth. In the latter case, we restart with a new random angle θ.

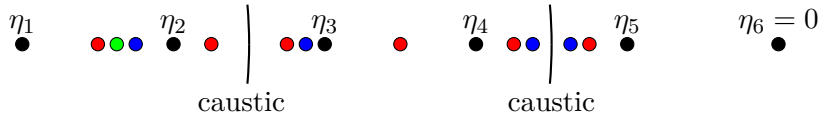


Figure 14: Greedy divide-and-conquer approach for constructing transport paths. Points of depth 0, 1, 2 and 3 (black, red, blue and green).

To check numerically whether a set S is a prediction set of $f^{-1}(\{\eta\})$, we apply the harmonic Newton method for $f - \eta$ to S. We stop the iteration when

$$\frac{|z_{k+1} - z_k|}{|z_{k+1}|} < \varepsilon_1 \quad \text{or} \quad |f(z_{k+1}) - \eta| < \varepsilon_1, \tag{5.5}$$

or after $n_{\max}$ iterations have been performed; see, e.g., [33, Sect. 25.5]. If the iteration converges numerically for all $s \in S$, and if the respective solutions are mutually distinct, i.e.,

$$|H^{\infty}_{f-\eta}(s_1) - H^{\infty}_{f-\eta}(s_2)| \geq \varepsilon_2, \tag{5.6}$$

for all $s_1, s_2 \in S$ with $s_1 \neq s_2$, then S is a prediction set of $f^{-1}(\{\eta\})$. In our numerical examples, we used $\varepsilon_1 = 10^{-14}$, $\varepsilon_2 = 10^{-10}$ and $n_{\max} = 50$.

Numerical examples

All computations in this section were performed in `MATLAB R2019b` on an `i7-7700 4 × 3.60 GHz` with `16 GB RAM`. Our numerical examples follow our previous ones in [74, Sect. 5]. For more examples regarding the number of Newton iterations, divided substeps and restarts with new angles, we refer the reader to [74, Tab. 1, 2 and 3].

We will observe three key features of the transport of images method in our examples.

1. It always terminates with the correct number of zeros.
2. It is highly accurate in terms of the residual.
3. It is significantly faster than other general-purpose root finders, which are not problem-adapted.

Let us start by briefly describing the syntax of our implementation. To compute all zeros of f, we first construct a `struct` of function handles of f and of its Wirtinger derivatives with `harmonicRat`. The naming is intended to indicate that harmonic mappings on $\widehat{\mathbb{C}}$ are of the form

$$f(z) = \frac{p_1(z)}{p_2(z)} + \overline{\left(\frac{p_3(z)}{p_4(z)}\right)} + \sum_{j=1}^{n} A_j \log|z - z_j|; \tag{5.7}$$

see our proof in [74, Prop. 2.1]. The function `harmonicRat` requires as input row vectors of the coefficients of the polynomials p_1, p_2, p_3, p_4, and of the A_j and z_j, respectively. To the resulting `struct` we apply our root-finding routine `tiroots`. As an example, consider the transcendental harmonic mapping

$$f(z) = z^2 + \overline{\left(\frac{2z+1}{z^2+z}\right)} + 2\log|z|; \tag{5.8}$$

cf. (4.3). Note that we have previously studied its critical curves and caustics (Figure 5 (right) and Figure 6 (right)), and its four zeros (Figure 7, second plot). Pasting the following code to the `MATLAB` command line

```
rng(1);
tic;
fun = harmonicRat([1 0 0], [1], [2 1], [1 1 0], [2], [0]);
zer = tiroots(fun)
res = max(abs(fun.f(zer)))
toc;
```

yields the output

```
zer =
  -0.8775 + 0.9278i
  -0.8775 - 0.9278i
  -1.4569 - 0.0000i
  -0.5872 + 0.0000i
res =
   8.8818e-16
Elapsed time is 0.056032 seconds.
```

We observe the following: (1) The transport of images method terminates with the correct number of zeros. Note that this number is not an input parameter of the method. We would also like to emphasize that the log term makes a symbolic computation of the zeros of (5.8) very difficult, for example `Mathematica` cannot determine all zeros of this function symbolically. (2) The computation is performed within a fraction of a second on a standard computer. (3) The results are very accurate in terms of the maximum residual $|f(z_j)|$ at the computed zeros z_j. The high accuracy is due to the use of Newton's method as corrector rather than the special form of the transport of images method. In particular, the transport of images method inherits all bounds on the error or the residual from Newton's method on $\mathbb{R}^2$; see, e.g., [23, Sect. 2.1.1 and 2.2.1] or [33, Ch. 25].

In order to illustrate that the transport of images method indeed computes *all* zeros numerically, we will apply it in the sequel to the following families of non-degenerate harmonic mappings for which the number of zeros is known analytically. The fact that the number of zeros is known for only a few functions underlines once more how difficult it is in general to compute all zeros of harmonic mappings.

Function	Zeros	Max. res.	Time
(5.9)	9	1.7764e-15	45 ms
(5.10)	22	2.4825e-16	67 ms
(5.11)	40	9.6069e-16	114 ms

Table 1: Computing the zeros with `tiroots` for (5.9) with $n = 3$, (5.10) with $n = 7$, $\rho = 0.7$, and (5.11) with $n = 8$, $\rho = 0.7$, $\varepsilon = 0.2$; cf. [74, Tab. 1].

1. Wilmshurst's harmonic polynomial [85, p. 2080]

$$f(z) = (z-1)^n + z^n + \overline{\mathrm{i}(z-1)^n - \mathrm{i}z^n}, \quad n \geq 1, \tag{5.9}$$

has exactly n^2 zeros. This is the maximum number of zeros of a harmonic polynomial of degree $n \geq 1$ which has finitely many zeros; see Section 3.1.

2. The harmonic mapping of Mao, Petters and Witt [53],

$$f(z) = z - \overline{\left(\frac{z^{n-1}}{z^n - \rho^n}\right)}, \quad n \geq 3, \quad \rho > 0, \tag{5.10}$$

has exactly $3n + 1$ zeros for $\rho < \rho_c$, where ρ_c is the *critical radius* which depends only on n; see [47, Prop. 2.1] for further details.

3. Rhie's harmonic mapping [67]

$$f(z) = z - \overline{\left((1-\varepsilon)\frac{z^{n-1}}{z^n - \rho^n} + \frac{\varepsilon}{z}\right)}, \quad n \geq 3, \quad \rho, \varepsilon > 0, \tag{5.11}$$

has exactly $5n$ zeros if $\rho < \rho_c$ and $\varepsilon < \varepsilon^*$, where ε^* depends on n and ρ; see [47, Thm. 2.2] for the details. This is the maximum number of zeros of $f(z) = z - \overline{r(z)}$ with a rational function r of degree $n+1 \geq 2$; see also Theorem 3.1.

Note that the functions (5.10) and (5.11) and their zeros are of particular interest in the context of gravitational lensing; see Section 1.2.

As above, we apply `tiroots` to the functions (5.9), (5.10) and (5.11) for certain parameters n, ρ and ε. The results are shown in Table 1. We observe again that our method terminates with the correct number of zeros and that the computations are very fast and highly accurate. Next, we illustrate the transport of images method using the functions from Table 1, closely following our previous presentation in [74, Sect. 5].

Figure 15 visualizes the transport of images method for Wilmshurst's polynomial (5.9) with $n = 3$. The first two plots display the transport path and a zoom-in close to the origin. The remaining plots show the basins of attraction of $H_{f-\eta_k}$, $k = 1, 2, \ldots, 9$. The initial phase is illustrated in the

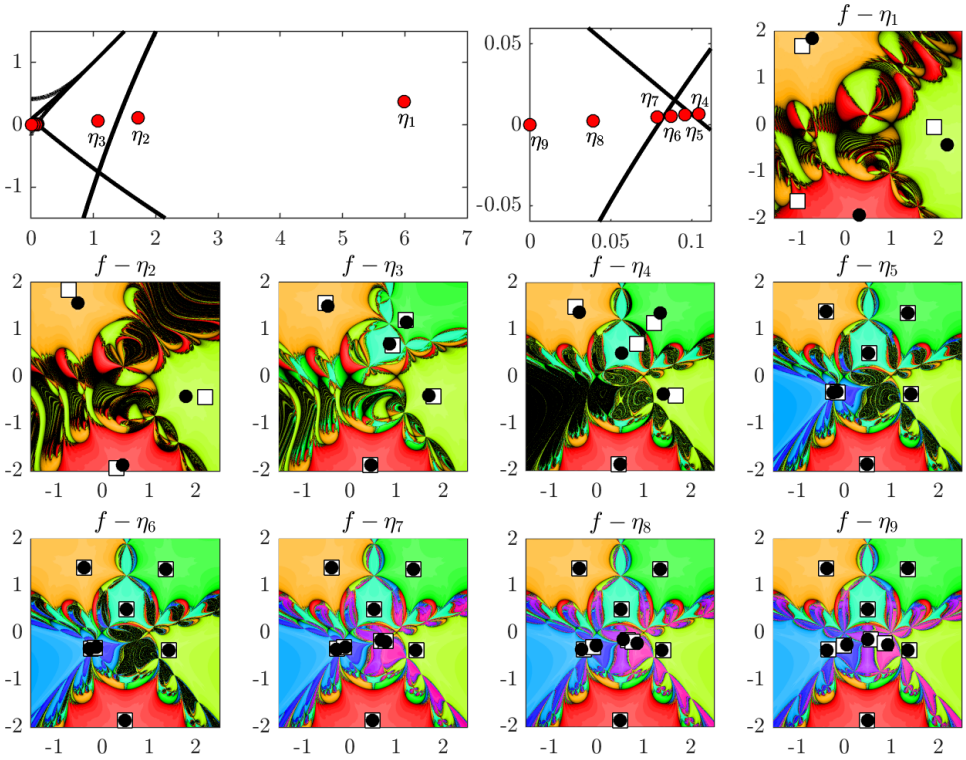


Figure 15: `tiroots` for (5.9) with $n = 3$. Transport path on $R_{\frac{\pi}{50}}$ (red dots), basins of attraction of $H_{f-\eta_k}$ with solutions of $f(z) = \eta_k$ (black dots) and corresponding prediction sets (white squares); see [74, Fig. 8].

upper right plot. Each solution of $f(z) = \eta_1$ (black dots) attracts one of the points constructed in the initial phase (white squares). Thus, these points form a prediction set of $f^{-1}(\{\eta_1\})$. The transport phase is visualized in the next eight plots. Again, the solutions of $f(z) = \eta_k$ (black dots) and the corresponding prediction sets (white squares) are shown. For each step, the prediction set of $f^{-1}(\{\eta_{k+1}\})$ consists of $f^{-1}(\{\eta_k\})$ (black dots in the previous plot) and, in case of a caustic crossing, additionally of the points $\zeta_\pm$ in (4.22). We see how the basins of attraction evolve while η_k "travels" along the transport path. New basins and their respective solutions appear in pairs immediately after a caustic crossing, i.e., for η_3 (green and cyan), η_5 (blue and light blue) and η_7 (pink and purple). Note that $f^{-1}(\{\eta_7\})$ is not a prediction set of $f^{-1}(\{\eta_9\})$. Therefore, the step from η_7 to η_9 was divided by introducing the point (of depth 1) $\eta_8 = (\eta_7 + \eta_9)/2$.

Figure 16 shows a transport path on $R_{\frac{\pi}{50}}$ for Rhie's function (5.11) with $n = 8$, $\rho = 0.7$ and $\varepsilon = 0.2$. Here, dividing the steps had to be done much more often. However, the total number of steps is still fairly small, which indicates that our greedy divide-and-conquer approach is at least competitive to adaptive step size control.

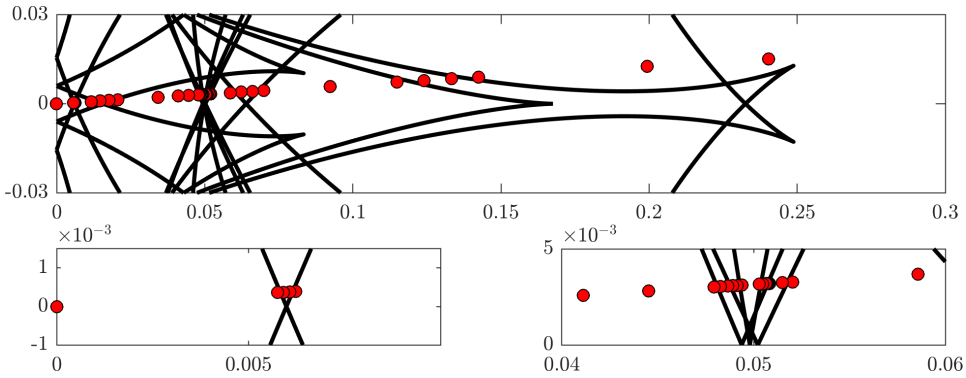


Figure 16: Transport path on $R_{\frac{\pi}{50}}$ for (5.11) with $n = 8$, $\rho = 0.7$ and $\varepsilon = 0.2$; cf. [74, Fig. 9].

As mentioned earlier, we are not aware of any method in the literature that is specialized to the numerical computation of *all* zeros for harmonic mappings. Nevertheless, in order to demonstrate the competitiveness of the transport of images method, we compare it with the `roots` command of Chebfun2[3], the state-of-the-art toolbox for numerical computations with (real or complex) smooth and bounded functions on a rectangle in the plane; see, e.g., the introductory article by Townsend and Trefethen [81]. Any method that can keep up with Chebfun2, even if it concerns only a single aspect for certain functions, is worth considering in our opinion. For us, Chebfun2 is the benchmark, although it is not specialized to harmonic mappings. In the `roots` command of Chebfun2, we can choose between two methods: `marchingsquares` and `resultant`; see [81, Sect. 7]. The latter is based on work by Nakatsukasa, Noferini and Townsend [59], and has also been implemented independently as `rootsb`[4].

The main idea in [59] for computing the common zeros of two real-valued smooth functions f and g, i.e., to solve $f(x, y) = g(x, y) = 0$, is the following. First, polynomial interpolants p and q of f and g are computed. Then the polynomial system $p(x, y) = q(x, y) = 0$ is solved using Bézout resultants. During the solution process, the interpolants are locally resampled if necessary. An important difference between `rootsb` and the implementation in Chebfun2 is that `rootsb` uses the original functions for resampling, while Chebfun2 uses global interpolants.

In the next example, we compare `tiroots`, Chebfun2 (`marchingsquares` and `resultant`) and `rootsb`. For $n = 11, 12, \ldots, 20$, we consider the function (5.10) with $\rho = 0.75$, and the function (5.11) with $\rho = 0.75$ and $\varepsilon = 0.25$. Since `rootsb` and Chebfun2 require bounded functions, we multiply (5.10)

[3] Chebfun2, version of December 16, 2021, `https://www.chebfun.org`.

[4] `rootsb`, version of December 16, 2021, `https://www.mathworks.com/matlabcentral/fileexchange/44084-computing-common-zeros-of-two-bivariate-functions`.

and (5.11) by their respective denominators to eliminate the poles, yielding

$$F_1(z) = z(\overline{z}^n - \rho^n) - \overline{z}^{n-1}, \tag{5.12}$$

$$F_2(z) = (|z|^2 - 1)\overline{z}^n + (\varepsilon - |z|^2)\rho^n. \tag{5.13}$$

We apply Chebfun2 and `rootsb` on the square $D = [-1.2, 1.2] \times [-1.2, 1.2]$ which contains all zeros of F_1 and F_2 for all n, ρ and ε as described above. Note that `tiroots` does not require any additional information about the location of the zeros in order to compute them. The number of computed zeros, the maximum residual $|f(z_j)|$ at the computed zeros z_j, and the timings for all methods are shown in Tables 2 and 3. We observe that `tiroots` is significantly faster than the other methods in both examples. Moreover, it terminates with the correct number of zeros for all functions and with a maximum residual that is close to machine precision. The same holds for `rootsb`. In contrast, Chebfun2 with `marching squares` and `resultant` computes too many zeros for some functions and we observe larger residuals. This is maybe due to the fact that the magnitudes of F_1 and F_2 vary highly on D, which is referred to as the *dynamic range issue*. For `resultant`, resampling with the original functions as in `rootsb` gives more accurate results, indeed, at the expense of speed. However, the slightly incorrect results should by no means be blamed on Chebfun2. They rather show how ill-behaved the considered functions are, and underline once again how difficult it is to compute their zeros.

A similar example, but for smaller n and averaged over a sample of the parameters ρ and ε, was given in [74, Fig. 11]. Moreover, in [74, Tab. 4] we considered functions with somewhat larger n. The `MATLAB` code implementing the input functions for the computations with `tiroots`, the `roots` command of Chebfun2 and `rootsb` reported in [74, Tab. 4] contained a small inaccuracy. This led us incorrectly conclude that `rootsb` computed not all zeros for the function (5.11) with the respective parameters n, ρ and ε. Here, we have therefore recomputed this example and show the results in Table 4. In fact, we observe again that `tiroots` and `rootsb` compute the zeros very accurately, but `tiroots` is significantly faster.

The results in Tables 1, 2, 3 and 4 can be reproduced using the `m`-files `/examples/diss_zur/tab_*.m`, which are contained in the `GitHub` repository of our `MATLAB` implementation[5].

Finally, we would like to mention that `tiroots` can also run into trouble even for small n. This may be due to several reasons. First of all, with our simplification of constructing transport paths only on rays R_θ, there is no longer a theoretical guarantee that such a path actually exists. Also, a potential problem is that the functions (5.10) and (5.11) have zeros which are nearly singular for $\rho \approx \rho_c$ or $\varepsilon \approx \varepsilon^*$. Another limitation of our implementation comes to light when the caustics are very nested. Then it would

[5] `https://github.com/transportofimages/Transport_of_images_Toolbox`

	tiroots			Chebfun2 (ms)			Chebfun2 (res)			rootsb		
n	zer.	res.	time	zer.	res.	time	zer.	res.	time	zer.	res.	time
11	34	5.0e-16	0.09 s	34	1.7e-13	1.37 s	34	1.7e-13	0.34 s	34	4.4e-16	0.68 s
12	37	6.7e-16	0.09 s	37	6.8e-14	1.62 s	37	7.3e-14	0.36 s	37	3.3e-16	0.99 s
13	40	7.1e-16	0.14 s	40	1.6e-13	1.80 s	40	1.4e-13	0.44 s	40	4.7e-16	1.10 s
14	43	8.2e-16	0.12 s	43	1.2e-13	2.17 s	43	3.7e-12	0.48 s	43	4.4e-16	1.76 s
15	46	7.9e-16	0.13 s	50	4.8e-13	2.23 s	46	1.9e-12	1.73 s	46	4.0e-16	2.50 s
16	49	8.6e-16	0.14 s	49	9.0e-13	2.61 s	49	8.3e-13	2.09 s	49	8.9e-16	2.69 s
17	52	9.0e-16	0.18 s	64	2.1e-12	2.80 s	52	3.1e-11	2.82 s	52	4.5e-16	4.85 s
18	55	8.5e-16	0.18 s	55	7.5e-12	3.04 s	55	6.8e-10	3.93 s	55	5.4e-16	6.27 s
19	58	5.2e-16	0.19 s	78	1.2e-10	3.47 s	58	4.5e-09	4.91 s	58	5.2e-16	6.90 s
20	61	9.0e-16	0.20 s	61	1.0e-11	3.52 s	61	8.8e-08	5.61 s	61	4.4e-16	9.31 s

Table 2: `tiroots`, Chebfun2 (`marching squares` (`ms`) and `resultant` (`res`)) and `rootsb` for (5.10) with $\rho = 0.75$. Number of computed zeros (zer.), maximum residual at the computed zeros (res.) and computation time.

	tiroots			Chebfun2 (ms)			Chebfun2 (res)			rootsb		
n	zer.	res.	time	zer.	res.	time	zer.	res.	time	zer.	res.	time
11	55	6.5e-16	0.16 s	55	9.2e-14	1.89 s	55	6.3e-13	0.72 s	55	3.4e-16	1.56 s
12	60	8.1e-16	0.14 s	60	1.3e-13	2.14 s	60	3.1e-12	0.93 s	60	4.8e-16	2.64 s
13	65	9.6e-16	0.33 s	65	3.8e-13	2.55 s	65	3.5e-10	1.37 s	65	4.0e-16	2.99 s
14	70	7.0e-16	0.18 s	70	8.8e-13	2.54 s	70	4.9e-10	2.65 s	70	3.5e-16	3.24 s
15	75	5.5e-16	0.20 s	75	2.0e-12	2.79 s	75	2.3e-09	4.07 s	75	4.7e-16	17.18 s
16	80	4.0e-16	0.21 s	80	7.5e-12	6.01 s	80	1.7e-09	4.53 s	80	4.0e-16	9.11 s
17	85	7.8e-16	0.24 s	85	5.5e-11	3.31 s	93	6.7e-08	5.44 s	85	4.6e-16	6.76 s
18	90	9.9e-16	0.27 s	90	1.2e-10	3.21 s	94	4.6e-08	6.68 s	90	5.5e-16	8.79 s
19	95	1.0e-15	0.27 s	95	6.3e-11	3.81 s	97	8.2e-08	6.69 s	95	2.4e-15	10.22 s
20	100	9.0e-16	0.49 s	100	3.4e-10	3.70 s	115	1.2e-07	9.64 s	100	6.8e-15	12.31 s

Table 3: `tiroots`, Chebfun2 (`marching squares` (`ms`) and `resultant` (`res`)) and `rootsb` for (5.11) with $\rho = 0.75$, $\varepsilon = 0.25$. Number of computed zeros (zer.), maximum residual at the computed zeros (res.) and computation time.

be beneficial to consider "nearby" caustic crossings together. This has not yet been implemented. However, very nested caustics also occur in some of the above examples, e.g., for (5.11) with $n = 20$, $\rho = 0.75$ and $\varepsilon = 0.25$. The caustics (including zoom-ins) and a phase plot of this function are displayed in Figure 17. Note that every ray R_θ intersects the caustics at least 39 times for this function. Despite the very nested caustics, `tiroots` computed the 100 zeros in half a second, as we see in the last row of Table 3.

	tiroots			Chebfun2			rootsb		
Function	zer.	res.	time	zer.	res.	time	zer.	res.	time
(5.10)	301	6.5e-16	5.27 s	8361	6.9e-00	36.33 s	301	1.1e-12	61.19 s
(5.11)	125	9.4e-16	1.60 s	161	8.3e-06	10.77 s	125	4.2e-16	11.51 s

Table 4: `tiroots`, the `roots` command of Chebfun2 and `rootsb` for (5.10) with $n = 100$, $\rho = 0.94$, and for (5.11) with $n = 25$, $\rho = 0.8$, $\varepsilon = 0.4$. Number of computed zeros (zer.), maximum residual at the computed zeros (res.) and computation time; cf. [74, Tab. 4].

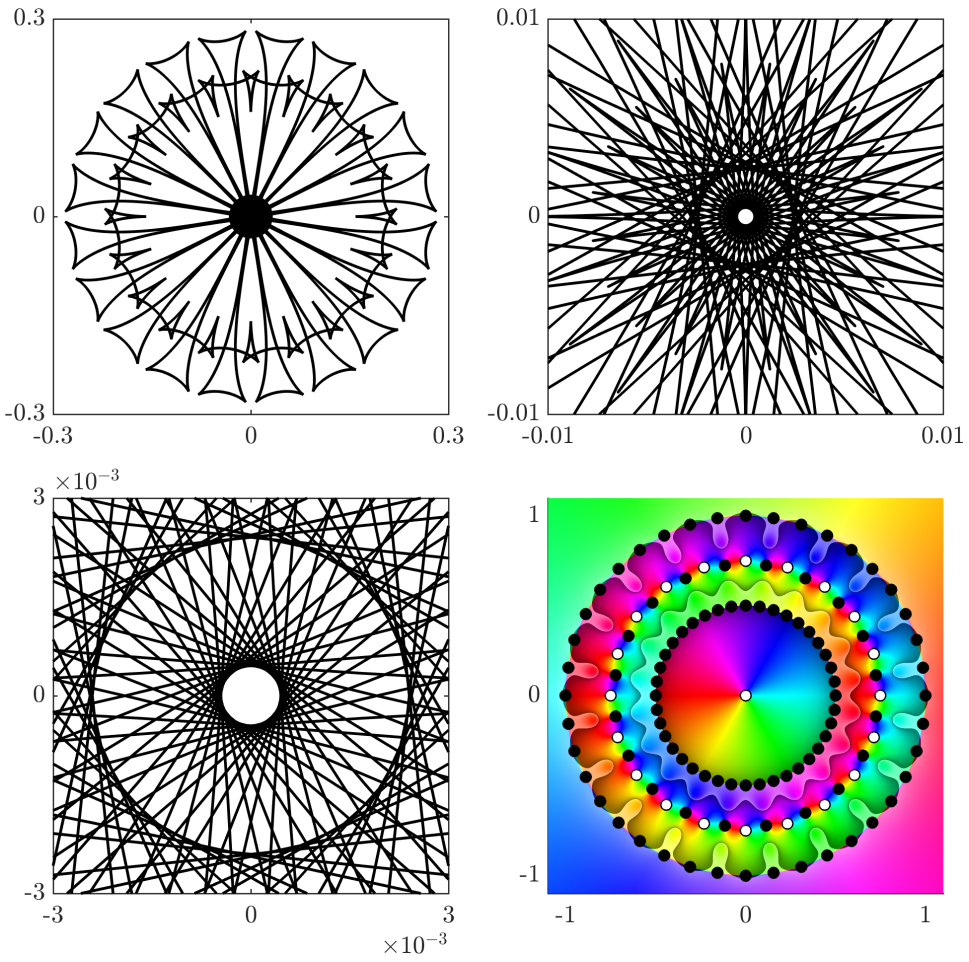


Figure 17: Caustics and a phase plot with zeros (black dots) and poles (white dots) of (5.11) with $n = 20$, $\rho = 0.75$ and $\varepsilon = 0.25$.

6 Conclusion and possible future research

In the present work, we studied the zeros of harmonic mappings f in the complex plane, and, more generally, the number of solutions of $f(z) = \eta$ for $\eta \in \mathbb{C}$. Our work was inspired and originally motivated by the theory of gravitational lensing, in which special cases of such functions and their zeros play an important role. In our studies, we used quite different mathematical tools, ranging from classical ones in complex analysis (the argument principle and dynamics of complex-valued functions) to numerical continuation methods. We contributed to the theory and numerics of harmonic mappings in the following three aspects.

In Section 3, we studied the number of zeros of rational harmonic functions and harmonic polynomials ([45], [73, Sect. 5]). For rational harmonic functions, we significantly sharpened bounds known from the previous literature on the maximum number of zeros, and we showed that any function attaining these bounds is non-singular. We also proved that for all $n > m \geq 1$ and every $k \in \{n, n+1, \ldots, V_{n,m}\}$ there exists a harmonic polynomial $f(z) = p(z) + \overline{q(z)}$ with $\deg(p) = n$, $\deg(q) = m$, and exactly k zeros.

In Section 4, we studied pre-images under (non-degenerate) harmonic mappings ([44, 72, 73]). We focused on how the number of pre-images changes with the parameter η. In doing so, we revealed an important connection between the number of pre-images of η and the position of η with respect to the caustics of f. This connection allows us to determine the number of pre-images of η from the winding number of the caustics about η, and even to derive the number geometrically from a plot of the caustics. While this has also been observed previously in the literature for certain special classes of harmonic mappings, a rigorous study, as we have carried out, had been lacking. Our novel results are a powerful tool to further advance the understanding of harmonic mappings. We used them to study the number of zeros of harmonic polynomials (Theorem 3.5) and to derive a numerical method for computing zeros of harmonic mappings.

In Section 5, we turned our attention to the numerical computation of zeros ([74]). Note that in almost all previous publications on the zeros of harmonic mappings (see, e.g., the references in Section 3.1), a lack of numerical examples can be observed. One reason probably is the fact that no method specialized to harmonic mappings was known in the literature. However, with the transport of images method, we introduced a first numerical method to compute all zeros of a non-singular and non-degenerate harmonic mapping. This makes the zeros of these functions easily accessible for practical computations. Our approach uses continuation and prediction-correction. Based on our results in Section 4, we proved that this method is theoretically guaranteed to terminate with all zeros after a finite number of continuation steps. Numerical examples with our `MATLAB` implementation

indicate that the transport of images method is highly accurate in terms of the residual, and significantly faster than the general-purpose root finders incorporated in Chebfun2.

Finally, we collect a few open questions and possible further research topics concerning the content of this thesis.

1. Geyer proved in [27, Cor. 2] that for every $n \geq 2$ there exists a harmonic polynomial $f(z) = p(z) - \overline{z}$ with exactly $3n - 2$ zeros; see Section 3.1. His proof relies on very advanced results in complex dynamics, but is not constructive. It is of interest to explicitly construct these *Geyer polynomials*.

2. While the bounds on the maximum number of zeros of rational harmonic functions in Theorem 3.1 are known to be sharp, the same has not yet been shown for our bounds in Theorem 3.2 (for $n_p > n_q + 1 > 1$ and for $n_p < n_q - 1$). However, we conjecture that all bounds in Theorem 3.2 are sharp.

3. It is of interest to further investigate the geometry of caustics for harmonic mappings, e.g., the number of cusps. Finding bounds on the number of cusps is also an important problem in gravitational lensing, e.g., for functions of the form (1.3); see [65, p. 1399]. The best known upper bound is quadratic in n (number of point masses), but our numerous experiments suggest that the maximum number is linear in n for functions as in (1.3).

4. Lyzzaik studied in [52] the valence of light harmonic mappings near their critical points. The *valence of f at z_0* is the maximum number of pre-images that an $\eta \in \mathbb{C}$ can have in any arbitrary small neighborhood of z_0; see [52, p. 136]. For critical points of the first and second kind, Lyzzaik determined the valence exactly, but for the critical points of the third kind, which are the points $z_0 \in \mathcal{C} \setminus \mathcal{M}$ with $\omega'(z_0) = 0$ in our terminology, he gave only an upper bound on the valence. It is of particular interest to complement Lyzzaik's study by giving exact bounds for this case as well. This would possibly also have implications for our next point.

5. Further investigating whether extremal harmonic mappings are non-singular is of interest. An immediate consequence of Theorem 4.6 is that if f is a harmonic mapping and $\eta \in f(\mathcal{C})$ is a *simple fold* with n pre-images, then there exists an $\widetilde{\eta} \in \mathbb{C} \setminus f(\mathcal{C})$ with at least $n + 1$ pre-images. We conjecture that this is also true for any kind of caustic point. In particular, this would imply that extremal harmonic mappings are always non-singular, which would generalize Theorem 3.4.

6. In Sections 4 and 5, we mainly considered harmonic mappings on $\widehat{\mathbb{C}}$. The extension of our results to functions on other domains $\Omega \subset \mathbb{C}$ is of interest. In that case, the *cluster set of f* (see, e.g., [7, p. 50]) must probably be treated as an "artificial" caustic.

7. In our theoretical results in Sections 4 and 5, we often used the imprecise and vague formulation "sufficiently small/large/close", e.g., in the Theorems 4.4, 4.5, 4.6, 5.2 and 5.3. Any estimate on the respective quantities is of interest, as it could improve our implementation of the transport of images method.

8. Another topic of possible future work ist the improvement of our implementation of the transport of images method, e.g., the correct treatment of multiple caustic crossings and the construction of transport paths that do not lie on rays. In addition, further adapting our method to the most relevant functions in the theory of gravitational lensing, e.g., functions as in (1.3), is also of great practical interest.

References

[1] D. Alexander and R. L. Devaney, *A century of complex dynamics*, in A century of advancing mathematics, Math. Assoc. America, Washington, DC, 2015, pp. 15–34.

[2] D. S. Alexander, *A history of complex dynamics. From Schröder to Fatou and Julia*, Aspects of Mathematics, Friedr. Vieweg & Sohn, Braunschweig, 1994.

[3] E. L. Allgower and K. Georg, *Introduction to numerical continuation methods*, vol. 45 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2003.

[4] J. H. An and N. W. Evans, *The Chang–Refsdal lens revisited*, Monthly Notices Roy. Astronom. Soc., 369 (2006), pp. 317–334.

[5] J. Arango, H. Arbeláez, and J. Rivera, *Orientation at singularities of harmonic functions*, Monatsh. Math., 193 (2020), pp. 737–762.

[6] V. I. Arnold, *Catastrophe theory*, Springer-Verlag, Berlin, third ed., 1992.

[7] M. B. Balk, *Polyanalytic Functions*, vol. 63 of Mathematical Research, Akademie-Verlag, Berlin, 1991.

[8] A. F. Beardon, *Iteration of rational functions. Complex analytic dynamical systems*, vol. 132 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1991.

[9] S. R. Bell, B. Ernst, S. Fancher, C. R. Keeton, A. Komanduru, and E. Lundberg, *Spiral galaxy lensing: a model with twist*, Math. Phys. Anal. Geom., 17 (2014), pp. 305–322.

[10] C. Bénéteau and N. Hudson, *A survey on the maximal number of solutions of equations related to gravitational lensing*, in Complex analysis and dynamical systems, Trends Math., Birkhäuser/Springer, Cham, 2018, pp. 23–38.

[11] W. Bergweiler and A. Eremenko, *On the number of solutions of a transcendental equation arising in the theory of gravitational lensing*, Comput. Methods Funct. Theory, 10 (2010), pp. 303–324.

[12] P. M. Bleher, Y. Homma, L. L. Ji, and R. K. W. Roeder, *Counting zeros of harmonic rational functions and its application to gravitational lensing*, Int. Math. Res. Not. IMRN, (2014), pp. 2245–2264.

[13] M. Brilleslyper, J. Brooks, M. Dorff, R. Howell, and L. Schaubroeck, *Zeros of a one-parameter family of harmonic trinomials*, Proc. Amer. Math. Soc. Ser. B, 7 (2020), pp. 82–90.

[14] D. Bshouty, W. Hengartner, and M. Naghibi-Beidokhti, *p-valent harmonic mappings with finite Blaschke dilatations*, Ann. Univ. Mariae Curie-Skłodowska Sect. A, 53 (1999), pp. 9–26.

[15] D. Bshouty, W. Hengartner, and T. Suez, *The exact bound on the number of zeros of harmonic polynomials*, J. Anal. Math., 67 (1995), pp. 207–218.

[16] D. Bshouty and A. Lyzzaik, *On Crofoot-Sarason's conjecture for harmonic polynomials*, Comput. Methods Funct. Theory, 4 (2004), pp. 35–41.

[17] D. Bshouty and A. Lyzzaik, *Problems and conjectures in planar harmonic mappings*, J. Anal., 18 (2010), pp. 69–81.

[18] L. Carleson and T. W. Gamelin, *Complex dynamics*, Universitext: Tracts in Mathematics, Springer-Verlag, New York, 1993.

[19] J. Clunie and T. Sheil-Small, *Harmonic univalent functions*, Ann. Acad. Sci. Fenn. Ser. A I Math., 9 (1984), pp. 3–25.

[20] A. B. Congdon and C. R. Keeton, *Principles of Gravitational Lensing - Light Deflection as a Probe of Astrophysics and Cosmology*, Springer Praxis Books, Springer, Cham, 2018.

[21] R. De Leo, *Conjectures about simple dynamics for some real Newton maps on* $\mathbb{R}^2$, Fractals, 27, Art. ID 19500993 (2019), pp. 1–22.

[22] R. De Leo, *Dynamics of Newton maps of quadratic polynomial maps of* $\mathbb{R}^2$ *into itself*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 30, Art. ID 2030027 (2020), pp. 1–21.

[23] P. Deuflhard, *Newton methods for nonlinear problems*, vol. 35 of Springer Series in Computational Mathematics, Springer, Heidelberg, 2011.

[24] P. Duren, *Harmonic mappings in the plane*, vol. 156 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2004.

[25] P. Duren, W. Hengartner, and R. S. Laugesen, *The argument principle for harmonic functions*, Amer. Math. Monthly, 103 (1996), pp. 411–415.

[26] C. D. Fassnacht, C. R. Keeton, and D. Khavinson, *Gravitational lensing by elliptical galaxies, and the Schwarz function*, in Analysis and mathematical physics, Trends Math., Birkhäuser, Basel, 2009, pp. 115–129.

[27] L. Geyer, *Sharp bounds for the valence of certain harmonic polynomials*, Proc. Amer. Math. Soc., 136 (2008), pp. 549–555.

[28] W. J. Gilbert, *Generalizations of Newton's method*, Fractals, 9 (2001), pp. 251–262.

[29] J. D. Hauenstein, A. Lerario, E. Lundberg, and D. Mehta, *Experiments on the zeros of harmonic polynomials using certified counting*, Exp. Math., 24 (2015), pp. 133–141.

[30] J. D. Hauenstein and A. J. Sommese, *What is numerical algebraic geometry?*, J. Symbolic Comput., 79 (2017), pp. 499–507.

[31] J. D. Hauenstein and F. Sottile, *Algorithm 921: alphaCertified: certifying solutions to polynomial systems*, ACM Trans. Math. Software, 38, Art. No. 28 (2012), pp. 1–20.

[32] W. Hengartner and G. Schober, *Univalent harmonic functions*, Trans. Amer. Math. Soc., 299 (1987), pp. 1–31.

[33] N. J. Higham, *Accuracy and stability of numerical algorithms*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 2002.

[34] J. Hubbard, D. Schleicher, and S. Sutherland, *How to find all roots of complex polynomials by Newton's method*, Invent. Math., 146 (2001), pp. 1–33.

[35] J. H. Hubbard and P. Papadopol, *Newton's method applied to two quadratic equations in $\mathbb{C}^2$ viewed as a global dynamical system*, Mem. Amer. Math. Soc., 191 (2008), pp. vi+146.

[36] D. Khavinson, S.-Y. Lee, and A. Saez, *Zeros of harmonic polynomials, critical lemniscates, and caustics*, Complex Anal. Synerg., 4, Art. No. 2 (2018), pp. 1–20.

[37] D. Khavinson and E. Lundberg, *Transcendental harmonic mappings and gravitational lensing by isothermal galaxies*, Complex Anal. Oper. Theory, 4 (2010), pp. 515–524.

[38] D. Khavinson and G. Neumann, *On the number of zeros of certain rational harmonic functions*, Proc. Amer. Math. Soc., 134 (2006), pp. 1077–1085.

[39] D. Khavinson and G. Neumann, *From the fundamental theorem of algebra to astrophysics: a "harmonious" path*, Notices Amer. Math. Soc., 55 (2008), pp. 666–675.

[40] D. Khavinson and G. Świątek, *On the number of zeros of certain harmonic polynomials*, Proc. Amer. Math. Soc., 131 (2003), pp. 409–414.

[41] L. Kuznia and E. Lundberg, *Fixed points of conjugated Blaschke products with applications to gravitational lensing*, Comput. Methods Funct. Theory, 9 (2009), pp. 435–442.

[42] S.-Y. Lee, A. Lerario, and E. Lundberg, *Remarks on Wilmshurst's theorem*, Indiana Univ. Math. J., 64 (2015), pp. 1153–1167.

[43] S.-Y. Lee and A. Saez, *A new lower bound for the maximal valence of harmonic polynomials*, Comput. Methods Funct. Theory, 17 (2017), pp. 139–149.

[44] J. Liesen and J. Zur, *How constant shifts affect the zeros of certain rational harmonic functions*, Comput. Methods Funct. Theory, 18 (2018), pp. 583–607.

[45] J. Liesen and J. Zur, *The maximum number of zeros of $r(z) - \overline{z}$ revisited*, Comput. Methods Funct. Theory, 18 (2018), pp. 463–472.

[46] R. Luce and O. Sète, *The index of singular zeros of harmonic mappings of anti-analytic degree one*, Complex Var. Elliptic Equ., 66 (2021), pp. 1–21.

[47] R. Luce, O. Sète, and J. Liesen, *Sharp parameter bounds for certain maximal point lenses*, Gen. Relativity Gravitation, 46, Art. No. 1736 (2014), pp. 1–16.

[48] R. Luce, O. Sète, and J. Liesen, *A note on the maximum number of zeros of $r(z) - \overline{z}$*, Comput. Methods Funct. Theory, 15 (2015), pp. 439–448.

[49] E. Lundberg, *The valence of harmonic polynomials viewed through the probabilistic lens*, arXiv:math/2201.00788v1, (2022).

[50] A. Lyzzaik, *On the valence of some classes of harmonic maps*, Math. Proc. Cambridge Philos. Soc., 110 (1991), pp. 313–325.

[51] A. Lyzzaik, *The geometry of some classes of folding polynomials*, Complex Variables Theory Appl., 20 (1992), pp. 145–155.

[52] A. Lyzzaik, *Local properties of light harmonic mappings*, Canad. J. Math., 44 (1992), pp. 135–153.

[53] S. Mao, A. O. Petters, and H. J. Witt, *Properties of point mass lenses on a regular polygon and the problem of maximum number of images*, in The Eighth Marcel Grossmann Meeting, Part A, B (Jerusalem, 1997), World Sci. Publ., River Edge, NJ, 1999, pp. 1494–1496.

[54] M. J. Martín, *Harmonic Schwarzian derivative and methods of approximation of zeros*, arXiv:math/2201.03009v1, (2022).

[55] J. Milnor, *Dynamics in one complex variable*, vol. 160 of Annals of Mathematics Studies, Princeton University Press, Princeton, NJ, third ed., 2006.

[56] A. Morgan, *Solving polynomial systems using continuation for engineering and scientific problems*, vol. 57 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009.

[57] A. P. Morgan, A. J. Sommese, and C. W. Wampler, *Computing singular solutions to nonlinear analytic systems*, Numer. Math., 58 (1991), pp. 669–684.

[58] A. P. Morgan, A. J. Sommese, and C. W. Wampler, *A power series method for computing singular solutions to nonlinear analytic systems*, Numer. Math., 63 (1992), pp. 391–409.

[59] Y. Nakatsukasa, V. Noferini, and A. Townsend, *Computing the common zeros of two bivariate functions via Bézout resultants*, Numer. Math., 129 (2015), pp. 181–209.

[60] G. Neumann, *Valence of complex-valued planar harmonic functions*, Trans. Amer. Math. Soc., 357 (2005), pp. 3133–3167.

[61] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York - London, 1970.

[62] H.-O. Peitgen, M. Prüfer, and K. Schmitt, *Global aspects of the continuous and discrete Newton method: a case study*, Acta Appl. Math., 13 (1988), pp. 123–202.

[63] H.-O. Peitgen and P. H. Richter, *The beauty of fractals. Images of complex dynamical systems*, Springer-Verlag, Berlin, 1986.

[64] R. Peretz and J. Schmid, *On the zero sets of certain complex polynomials*, in Proceedings of the Ashkelon Workshop on Complex Function Theory (1996), vol. 11 of Israel Math. Conf. Proc., Bar-Ilan Univ., Ramat Gan, 1997, pp. 203–208.

[65] A. O. Petters, *Gravity's action on light*, Notices Amer. Math. Soc., 57 (2010), pp. 1392–1409.

[66] A. O. Petters, H. Levine, and J. Wambsganss, *Singularity theory and gravitational lensing*, vol. 21 of Progress in Mathematical Physics, Birkhäuser Boston, Inc., Boston, MA, 2001.

[67] S. H. Rhie, *n-point gravitational lenses with 5(n-1) images*, arXiv:astro-ph/0305166v1, (2003).

[68] J. Roe, *Winding around.*, vol. 76 of Student Mathematical Library, American Mathematical Society, Providence, RI; Mathematics Advanced Study Semesters, University Park, PA, 2015.

[69] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses*, Springer-Verlag, Berlin - Heidelberg, 1992.

[70] O. Sète, R. Luce, and J. Liesen, *Creating images by adding masses to gravitational point lenses*, Gen. Relativity Gravitation, 47, Art. No. 42 (2015), pp. 1–8.

[71] O. Sète, R. Luce, and J. Liesen, *Perturbing rational harmonic functions by poles*, Comput. Methods Funct. Theory, 15 (2015), pp. 9–35.

[72] O. Sète and J. Zur, *A Newton method for harmonic mappings in the plane*, IMA J. Numer. Anal., 40 (2020), pp. 2777–2801.

[73] O. Sète and J. Zur, *Number and location of pre-images under harmonic mappings in the plane*, Ann. Fenn. Math., 46 (2021), pp. 225–247.

[74] O. Sète and J. Zur, *The transport of images method: computing all zeros of harmonic mappings by continuation*, IMA J. Numer. Anal., (2021). Online first publication.

[75] T. Sheil-Small, *Three unsolved problems on polynomials*, in Tagungsbericht 7/1992 - Workshop 9208 - Funktionentheorie, Mathematisches Forschungsinstitut Oberwolfach, 1992.

[76] T. Sheil-Small, *Complex Polynomials*, vol. 75 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2002.

[77] S. Smale, *Newton's method estimates from data at one point*, in The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985), Springer, New York, 1986, pp. 185–196.

[78] A. J. Sommese, J. Verschelde, and C. W. Wampler, *Introduction to numerical algebraic geometry*, in Solving polynomial equations, vol. 14 of Algorithms Comput. Math., Springer, Berlin, 2005, pp. 301–335.

[79] A. J. Sommese and C. W. Wampler, II, *The numerical solution of systems of polynomials*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.

[80] T. J. Suffridge and J. W. Thompson, *Local behavior of harmonic mappings*, Complex Variables Theory Appl., 41 (2000), pp. 63–80.

[81] A. Townsend and L. N. Trefethen, *An extension of Chebfun to two dimensions*, SIAM J. Sci. Comput., 35 (2013), pp. C495–C518.

[82] J. L. Varona, *Graphic and numerical comparison between iterative methods.*, Math. Intell., 24 (2002), pp. 37–46.

[83] E. Wegert, *Visual complex functions. An introduction with phase portraits*, Birkhäuser/Springer Basel AG, Basel, 2012.

[84] E. Wegert and G. Semmler, *Phase plots of complex functions: a journey in illustration*, Notices Amer. Math. Soc., 58 (2011), pp. 768–780.

[85] A. S. Wilmshurst, *The valence of harmonic polynomials*, Proc. Amer. Math. Soc., 126 (1998), pp. 2077–2081.

How constant shifts affect the zeros of certain rational harmonic functions*

Jörg Liesen† Jan Zur†

Abstract

We study the effect of constant shifts on the zeros of rational harmomic functions $f(z) = r(z) - \overline{z}$. In particular, we characterize how shifting through the caustics of f changes the number of zeros and their respective orientations. This also yields insight into the nature of the singular zeros of f. Our results have applications in gravitational lensing theory, where certain such functions f represent gravitational point-mass lenses, and a constant shift can be interpreted as the position of the light source of the lens.

Keywords: Rational harmonic functions, critical curves and caustics, cusp and fold points, singular zeros, gravitational lensing.

Mathematics subject classification (2010): 30D05; 31A05; 85A04.

1 Introduction

The number and location of the zeros of rational harmonic functions of the form

$$f(z) = r(z) - \overline{z}, \tag{1.1}$$

where r is a rational function, have been intensively studied in recent years. An important result of Khavinson and Neumann [5] says that if $\deg(r) \geq 2$, then f may have at most $5\deg(r) - 5$ zeros. As shown by a construction of Rhie [17], this bound on the maximal number of zeros is sharp in the sense that for every $n \geq 2$ there exists a rational harmonic function as in (1.1) with $n = \deg(r)$ and exactly $5n - 5$ zeros. Several authors have derived more refined bounds on the maximal number of zeros which depend on the

*A final version of this preprint manuscript appeared as follows: J. Liesen and J. Zur. How constant shifts affect the zeros of certain rational harmonic functions. *Comput. Methods Funct. Theory* (2018), 18(4), pp. 583–607, doi:10.1007/s40315-018-0240-8.

†TU Berlin, Institute of Mathematics, MA 3-3, Straße des 17. Juni 136, 10623 Berlin, Germany. {liesen,zur}@math.tu-berlin.de

degrees of the numerator and denominator polynomials of r; see, e.g., [8] and the references given there.

Rhie made her construction in the context of astrophysics, where certain rational harmonic functions model gravitational lenses based on n point-masses; see the Introduction of [21] for a brief summary of Rhie's construction, and [10] for a detailed analysis. Descriptions of the connection between complex analysis and gravitational lensing are given, for example, in the articles [2, 6, 13, 15], and a comprehensive treatment can be found in the monographs [14, 18]. The function modeling the gravitational point-mass lens is a special case of (1.1), namely

$$f(z) = \overline{z} - r(z), \quad \text{where} \quad r(z) = \sum_{k=1}^{n} \frac{m_k}{z - z_k}. \tag{1.2}$$

The poles $z_1, \ldots, z_n \in \mathbb{C}$ represent the position of the respective point-masses $m_1, \ldots, m_n > 0$ in the lens plane. For a fixed $\eta \in \mathbb{C}$, a solution of $f(z) = \eta$, or equivalently a zero of $f_\eta(z) = f(z) - \eta$, represents a lensed image of a light source at the position η in the source plane. Of great importance in this application is the behavior of the zeros under movements of the light source, i.e., changes of the parameter η. Using explicit computations, Schneider and Weiss studied this behavior for two point-masses, i.e., $n = 2$ in (1.2), in their frequently cited paper [19]. The same model was analyzed extensively by Witt and Petters [24]. Schneider, Ehlers and Falco pointed out in [18, p. 265], that the two point-mass lens is already fairly complicated to analyze in detail. Petters, Levine and Wambsganss gave a more general analysis in [14, Part III] based on the Taylor series of the gravitational lens potential associated with the lensing map $z \mapsto f_\eta(z)$. By truncating the Taylor series and neglecting higher order terms, they obtained an approximation to the lensing map's local quantitative behavior in [14, Section 9.2].

In this paper we give a rigorous analysis of the effect of varying the parameter η on the zeros of rational harmonic functions of the form

$$f_\eta(z) = f(z) - \eta, \quad \text{where } f \text{ is as in (1.1).}$$

In particular, we study the behavior of the zeros when η crosses a caustic of f (see Section 2 for a definition of this term). Apart from advancing the overall understanding of rational harmonic functions, our goal is to confirm and generalize the above mentioned results published in the astrophysics literature. One of the consequences of our findings is that may change the number of zeros of f_η by $4\deg(r) - 6$. Thus, the effect of varying η is considerably different from the effect of perturbing f by poles that was studied in [21].

The paper is organized as follows. In Section 2 we discuss the mathematical background, in particular the critical curves, caustics, and exceptional points (zeros and poles) of f. In Section 3 we focus on constant shifts that do

not affect the number of zeros. Our main results are contained in Section 4, where we study in detail how shifting η across a caustic of f affects the zeros. Here we distinguish between shifting through fold and cusp points of f. Our results on shifts also yield some insight into the nature of the singular zeros of f. In Section 5 we give examples that illustrate our results and a brief outlook on possible extensions and further work in this area.

2 Critical curves, caustics and the Poincaré index

Let a rational harmonic function $f(z) = r(z) - \overline{z}$ with $\deg(r) \geq 2$ be given. Using the Wirtinger derivatives ∂_z and $\partial_{\overline{z}}$ we can write the *Jacobian* of f as

$$J_f(z) = |\partial_z f(z)|^2 - |\partial_{\overline{z}} f(z)|^2 = |r'(z)|^2 - 1.$$

The points $z \in \mathbb{C}$ where J_f vanishes, i.e., where $|r'(z)| = 1$, are called the *critical points* of f. We denote the set of the critical points by $\mathcal{C}$. The critical points of f are the preimages of the unit circle $|w| = 1$ under the map $w = r'(z)$, which is analytic (and non-constant) in $\mathbb{C}$, except at the finitely many poles of $r'(z)$. Thus, the critical points form finitely many closed curves that separate the complex plane into regions where $J_f(z) > 0$ and hence f is *sense-preserving*, and $J_f(z) < 0$ and hence f is *sense-reversing*. We denote these regions by Ω_+ and Ω_-, respectively, so that we have the disjoint partitioning $\mathbb{C} = \mathcal{C} \cup \Omega_+ \cup \Omega_-$. Each closed curve in the set $\mathcal{C}$ is called a *critical curve* of f.

The necessary condition for a stationary point of the Jacobian of f is

$$\partial_{\overline{z}} J_f(z) = r'(z)\overline{r''(z)} = 0,$$

and hence the condition $r''(z) \neq 0$ for all $z \in \mathcal{C}$ implies that no critical point of f is a saddle-point of J_f. Then the critical curves of f are smooth Jordan curves, and in particular they do not intersect each other; see the left plot in Figure 1 for an example. A function f with this property is called *non-degenerate*, and in the following we will always assume that the given f is such a function.

In this case the critical curves yield a disjoint partitioning of $\Omega_+ \cup \Omega_-$ into finitely many open and connected subsets $A_1, \ldots, A_m$, where $\partial A_j \subseteq \mathcal{C}$, and either $A_j \subseteq \Omega_+$ or $A_j \subseteq \Omega_-$, for $j = 1, \ldots, m$, and we write

$$\mathcal{A} := \{A_1, \ldots, A_m\}. \tag{2.1}$$

Exactly one of the sets A_j is unbounded, and we sometimes denote this set by A_∞. On the two bordered sets of a given critical curve is f always differently oriented. This is a consequence of the maximum modulus principle applied to the functions r' and $1/r'$.

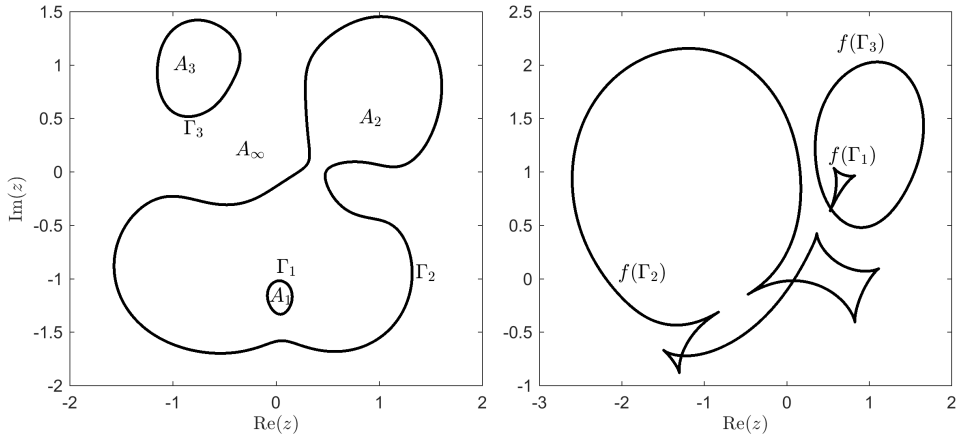


Figure 1: Critical curves Γ_i and $\mathcal{A}$ (left), caustics $f(\Gamma_i)$ (right)

The elements of the set $f(\mathcal{C})$ are called the *caustic points* of f, and for each critical curve Γ, the curve $f(\Gamma)$ is called a *caustic* of f. Unlike a critical curve, a caustic of f may intersect itself as well as other caustics of f, and a caustic of f need not be smooth; see the right plot of Figure 1 for examples.

The singularities on a caustic of f are called *cusp points*, and all other caustic points of f are called *fold points*; cf. [14, p. 88]. In order to characterize a cusp point, note that the unique tangent at a critical point $z_0 \in \mathcal{C}$ is given by

$$g(t) = ht + z_0, \quad h := \frac{ir'(z_0)\overline{r''(z_0)}}{|r'(z_0)\overline{r''(z_0)}|}, \tag{2.2}$$

where we use that the gradient of the Jacobian is orthogonal with respect to its contour line, i.e., the critical curve, and where the normalization of the direction h will be convenient in our derivations in Section 4. The linearization of f at $z_0 \in \mathcal{C}$ has the form

$$L_{z_0}(z) := r(z_0) + e^{i\varphi}(z - z_0) - \overline{z},$$

where we use that $r'(z_0) = e^{i\varphi}$ for some $\varphi \in [0, 2\pi)$. After some small manipulations we obtain

$$L_{z_0}(g(t)) = 2ie^{i\varphi/2}\,\mathrm{Im}(e^{i\varphi/2}h)t + r(z_0) - \overline{z}_0,$$

which shows that the tangent direction at the caustic point $f(z_0) \in f(\mathcal{C})$ is given by $ie^{i\varphi/2}$. Moreover, L_{z_0} maps the tangent at the critical curve to a single point if and only if

$$\mathrm{Im}(e^{i\varphi/2}h) = \mathrm{Im}(ie^{3i\varphi/2}\overline{r''(z_0)}) = 0,$$

or, equivalently,

$$\operatorname{Re}\left(\frac{r''(z_0)}{r'(z_0)^{3/2}}\right) = 0 \quad \Longleftrightarrow \quad \operatorname{Re}\left(\frac{r'(z_0)^{3/2}}{r''(z_0)}\right) = 0, \tag{2.3}$$

where the equivalence is defined since we assume that f is non-degenerate. Let us summarize these considerations.

Lemma 2.1. *Let $z_0 \in \mathcal{C}$ be a critical point of $f(z) = r(z) - \overline{z}$. Then the caustic point $f(z_0) \in f(\mathcal{C})$ is a cusp point if and only if* (2.3) *holds. (Each other caustic point $f(z_0) \in f(\mathcal{C})$ is called a fold point.)*

Petters and Witt [16] showed that if r is as in (1.2), then there can be at most $12n(n-1)$ cusp points; see also [14, Section 15.3.3]. The determination of a sharp upper bound on the number of cusp points was mentioned as an open research problem in [13]. The relation between the number of cusp points and the number of zeros for harmonic *polynomials* was recently studied in [4].

Now let $z_0 \in \mathbb{C}$ be such that $f(z_0) = r(z_0) - \overline{z}_0 = 0$. We call z_0 a *sense-preserving*, *sense-reversing*, or *singular zero* of f, if z_0 is an element of Ω_+, Ω_-, or $\mathcal{C}$, respectively. Note that if z_0 is a sense-preserving or sense-reversing zero, then there exists an $\varepsilon > 0$ such that f is sense-preserving or sense-reversing, respectively, on $B_\varepsilon(z_0)$, the open disk around z_0 with radius ε. The sense-preserving and sense-reversing zeros of f are also called the *regular zeros* of f. If f has only such zeros, f is called *regular*, and otherwise f is called *singular*.

We have the following simple but important relation between singular zeros and caustic points.

Proposition 2.2. *Let $f(z) = r(z) - \overline{z}$ with $\deg(r) \geq 2$ and $\eta \in \mathbb{C}$ be given. Then $f_\eta(z) := f(z) - \eta$ has a singular zero if and only if η is a caustic point of f.*

Proof. If $z_0 \in \mathbb{C}$ is a singular zero of f_η, then $z_0 \in \mathcal{C}$ and $f(z_0) - \eta = 0$, or $f(z_0) = \eta$, which means that η is a caustic point of f. On the other hand, if $\eta \in \mathbb{C}$ is a caustic point of f, then $f(z_0) = \eta$ for some $z_0 \in \mathcal{C}$, which means that z_0 is a singular zero of f_η. □

Sometimes we will use the contraposition of the statement of Proposition 2.2: If $\eta \in \mathbb{C}$ is *not* on a caustic of f, i.e., $\eta \notin f(\mathcal{C})$, then the shifted function $f_\eta(z) = f(z) - \eta$ does *not* have a singular zero and hence is regular.

Let us briefly recall the argument principle for continuous functions; see [1, Corollary 2.6], [22, Theorem 2.2], or [21, Section 2] for more details. Let Γ be a closed Jordan curve, and let f be a function that is continuous and nonzero on Γ. Then the *winding* of f on Γ is defined as the change in

the argument of $f(z)$ as z travels once around Γ in the positive direction, divided by 2π, i.e.,

$$V(f;\Gamma) := \frac{1}{2\pi}\Delta_\Gamma \arg f(z).$$

A point $z_0 \in \mathbb{C}$ is called an *exceptional point* of a function f, if f is either zero, not continuous, or not defined at z_0. If f is continuous and nonzero in a punctured neighborhood D of an exceptional point z_0, and hence the exceptional point z_0 is isolated, then the *Poincaré index* of f at z_0 is defined as $\operatorname{ind}(f;z_0) := V(f;\Gamma)$, where Γ is an arbitrary closed Jordan curve in D and around z_0. This can be seen as a generalization of the order of a zero or a pole of a meromorphic function; cf. [21, Example 2.5]. The Poincaré index is independent of the choice of the Jordan curve Γ, as long as z_0 is the only exceptional point of f in $\operatorname{int}(\Gamma)$, the interior of Γ. If there are several (isolated) exceptional points in $\operatorname{int}(\Gamma)$, we have the following theorem.

Theorem 2.3. *If Γ is a closed Jordan curve and the function f is continuous and nonzero on $\overline{\operatorname{int}(\Gamma)}$ except for finitely many exceptional points $z_1, \ldots, z_k \in \operatorname{int}(\Gamma)$, then*

$$V(f;\Gamma) = \sum_{j=1}^{k} \operatorname{ind}(f;z_j).$$

For the functions of our interest, which are continuous in $\mathbb{C}$ except for finitely many exceptional points, we have the following Poincaré indices; see [21, Proposition 2.7].

Proposition 2.4. *Let $f(z) = r(z) - \overline{z}$ with $\deg(r) \geq 2$ be given. The Poincaré index of f at a sense-preserving zero is $+1$, and at a sense-reversing it is -1. If z_0 is a pole of r of order m, then f is sense-preserving in a neighborhood of z_0, and the Poincaré index of f at z_0 is $-m$.*

The determination of the Poincaré index of a singular zero is more challenging. For the functions of our interest it may be -1, 0, or $+1$ (see Corollary 4.6 and its discussion), while for a general harmonic function it may even be undefined; see [3, p. 413].

The next result, which is an immediate consequence of Theorem 2.3 and Proposition 2.4, shows how we can use the argument principle in order to determine the number of zeros.

Corollary 2.5. *Let $f(z) = r(z) - \overline{z}$ with $\deg(r) \geq 2$ be given. If f is nonzero on a closed Jordan curve Γ and has no singular zero in $\operatorname{int}(\Gamma)$, then*

$$V(f;\Gamma) = N_+(f;\operatorname{int}(\Gamma)) - N_-(f;\operatorname{int}(\Gamma)) - P(f;\operatorname{int}(\Gamma)),$$

where $N_{(+,-)}(f;\operatorname{int}(\Gamma))$ denotes the number of sense-preserving and sense-reversing zeros, and $P(f;\operatorname{int}(\Gamma))$ denotes the number of poles (with multiplicities) of f in $\operatorname{int}(\Gamma)$.

Finally, we state a version of Rouché's theorem which we will frequently use in order to decide whether two functions have the same winding on a given Jordan curve. A short proof of this result is given [21, Theorem 2.3].

Theorem 2.6. *Let Γ be a closed Jordan curve and suppose that $f, g : \Gamma \to \mathbb{C}$ are continuous. If $|f(z) - g(z)| < |f(z)| + |g(z)|$ holds for all $z \in \Gamma$, then $V(f;\Gamma) = V(g;\Gamma)$.*

3 Constant shifts that do not affect the number of zeros

In this section we will begin our study of the effect of constant shifts on the zeros of a given *non-degenerate* rational harmonic function

$$f(z) = r(z) - \overline{z} \quad \text{with} \quad \deg(r) \geq 2 \quad \text{and} \quad \lim_{|z|\to\infty} |f(z)| = \infty. \tag{3.1}$$

As mentioned in [21, Remark 3.2], the assumption $\lim_{|z|\to\infty} |f(z)| = \infty$ is not restrictive. It only excludes functions with $r(z) = \alpha z + p(z)/q(z)$, where $|\alpha| = 1$ and $\deg(p) \leq \deg(q)$.

In addition to the notation established in Corollary 2.5, we denote by $N(f;A)$ the number of zeros of f in the set $A \subseteq \mathbb{C}$, and write $N(f) := N(f;\mathbb{C})$ for brevity. Moreover, by $N_s(f)$ we denote the number of singular zeros of f.

Our first result characterizes the zeros of the shifted function $f_\eta(z) = f(z) - \eta$ for a *sufficiently large* (real) shift $\eta > 0$.

Theorem 3.1. *Let f be as in (3.1) with $r(z) = \frac{p(z)}{q(z)}$, let $v_1, \dots, v_m$ be the poles of f with their respective multiplicities $\mu_1, \dots, \mu_m$, and let $k := \max(\deg(p) - \deg(q), 1)$. If $\eta > 0$ is sufficiently large, then f_η has exactly $\deg(q) + k$ zeros $z_1, \dots z_{\deg(q)+k}$, and it exists an $\varepsilon > 0$ such that*

(i) $N(f_\eta; B_\varepsilon(v_j)) = N_+(f_\eta; B_\varepsilon(v_j)) = \mu_j \quad$ *for* $j = 1, \dots, m$,

(ii) $z_j \in A_\infty \setminus \bigcup_{\ell=1}^m B_\varepsilon(v_\ell) \quad$ *for* $j = \deg(q) + 1, \dots, \deg(q) + k$.

Proof. In order to explain the general idea of the proof, assume that we are given some $\eta \gg 1$. Then $f_\eta(z_j) = 0$ means that $r(z_j) - \overline{z}_j \gg 1$, which can happen when the zero z_j of f_η is close to a pole of f, or when $|z_j| \gg 1$. These cases correspond to (i) and (ii), and we will now first prove the existence of the zeros in (i), and then of the additional zeros in (ii).

Case 1 (zeros close to a pole): In the neighborhood of any pole v_j of f we have $|r'(z)| \gg 1$, and hence f is sense-preserving. Therefore we can find an $\varepsilon > 0$ such that

(a) f is sense-preserving on $B_\varepsilon(v_j)$ for all $j = 1, \dots, m$,

(b) $B_\varepsilon(z_j) \cap B_\varepsilon(z_\ell) = \varnothing$ for all $j, \ell \in \{1, \dots, m\}$ with $j \neq \ell$.

Now consider any $\eta \geq 2\max\{|f(z)| : z \in \bigcup_{j=1}^{m} \partial B_\varepsilon(v_j)\} > 0$, and let $g(z) \equiv -\eta$. Then

$$|f_\eta(z) - g(z)| = |f(z)| < |\eta| \leq |f_\eta(z)| + |g(z)| \quad \text{for all } z \in \bigcup_{j=1}^{m} \partial B_\varepsilon(v_j).$$

Since $V(g; \partial B_\varepsilon(v_j)) = 0$, the function f has a pole of order μ_j in $B_\varepsilon(v_j)$, and f is sense-preserving in $B_\varepsilon(v_j)$, Theorem 2.6 and Proposition 2.4 yield

$$N(f_\eta; B_\varepsilon(v_j)) = N_+(f_\eta; B_\varepsilon(v_j)) = \mu_j \quad \text{for all } j = 1, \ldots, m,$$

which proves the existence of the zeros $z_1, \ldots, z_{\deg(q)}$ as stated in (i).

Case 2 (zeros away from the poles): We need to distinguish four cases according to the degrees of p and q.

(a) $\deg(p) < \deg(q)$, hence $k = 1$: In this case $\lim_{|z|\to\infty} |r(z)| = 0$. Therefore, if $\eta > 0$ is chosen large enough, there exists a $\delta > 0$, such that $B_\delta(-\eta) \subset A_\infty$, and we have $|r(z)| < \delta$ for all $z \in \partial B_\delta(-\eta)$, as well as $B_\delta(-\eta) \cap B_\varepsilon(v_j) = \varnothing$ for all $j = 1, \ldots, m$. Using the function $g(z) := -\overline{z} - \eta$, which has $-\eta$ as its only zero, we obtain

$$|f_\eta(z) - g(z)| = |r(z)| < \delta = |g(z)| \leq |f_\eta(z)| + |g(z)| \text{ for all } z \in \partial B_\delta(-\eta).$$

Using Theorem 2.6 and Proposition 2.4 we get

$$V(f_\eta; \partial B_\delta(-\eta)) = V(g; \partial B_\delta(-\eta)) = -1,$$

since in our region of interest $|r'(z)| < 1$, and therefore $N(f_\eta; B_\delta(-\eta)) = N(g; B_\delta(-\eta)) = 1$. This shows the existence of one additional zero of f_η, which is contained in the set A_∞.

(b) $\deg(p) = \deg(q)$, hence $k = 1$: In this case we have $p = cq + \widetilde{q}$ for some (nonzero) $c \in \mathbb{C}$ and some polynomial $\widetilde{q}$ with $\deg(\widetilde{q}) < \deg(q)$. Hence $r(z) = c + \widetilde{q}(z)/q(z)$, where $\lim_{z\to\infty} |\widetilde{q}(z)/q(z)| = 0$. We can now apply the same argument as in the previous case with the function $g(z) := -\overline{z} - \eta + c$, using the disk $B_\delta(-\eta + \overline{c})$ for sufficiently small $\delta > 0$.

In the next two cases we will use that whenever $\deg(p) - \deg(q) > 0$, we can write

$$r(z) = cz^k + \rho(z) + \widetilde{r}(z), \quad k = \deg(p) - \deg(q), \tag{3.2}$$

for some (nonzero) $c \in \mathbb{C}$, some polynomial ρ of degree at most $k - 1$, and some rational function $\widetilde{r}$ with $\lim_{|z|\to\infty} |\widetilde{r}(z)| = 0$.

(c) $\deg(p) = \deg(q) + 1$, hence $k = 1$: Our general assumption $\lim_{|z|\to\infty} |f(z)| = \infty$ implies that in this case we have (3.2) with $|c| \neq 1$. We will first show that for each η the function $g(z) := cz - \eta - \overline{z}$ has exactly

one zero. Writing $c = \mathrm{Re}(c) + i\,\mathrm{Im}(c)$ and $z = x + iy$, the equation $g(z) = 0$ can be written as

$$\begin{bmatrix} \mathrm{Re}(c) - 1 & -\mathrm{Im}(c) \\ \mathrm{Im}(c) & \mathrm{Re}(c) + 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \eta \\ 0 \end{bmatrix}.$$

The determinant of the matrix is $\mathrm{Re}(c)^2 - 1 + \mathrm{Im}(c)^2 = |c|^2 - 1 \neq 0$. Denoting the unique zero of g by z_g and using that $\lim_{|z|\to\infty} |\widetilde{r}(z)| = 0$, we can choose $\eta > 0$ sufficiently large so that $|\widetilde{r}(z)| < |cz - \eta - \overline{z}|$ holds for some $\delta > 0$ and all $z \in \partial B_\delta(z_g)$. As above we can assume that $B_\delta(z_g) \subset A_\infty$ and that $B_\delta(z_g) \cap B_\varepsilon(v_j) = \varnothing$ for all $j = 1, \ldots, m$. We then get

$$|f_\eta(z) - g(z)| = |\widetilde{r}(z)| < |cz - \eta - \overline{z}| = |g(z)| \leq |f_\eta(z)| + |g(z)|$$

for all $z \in \partial B_\delta(z_g)$, so that $N(f_\eta; B_\delta(z_g)) = N(g; B_\delta(z_g)) = 1$ follows from Theorem 2.6 and Proposition 2.4.

(d) $\deg(p) > \deg(q) + 1$, hence $k > 1$: For a given $\eta > 0$, let $\eta_1, \ldots, \eta_k$ be the $k \geq 2$ zeros of $g(z) := cz^k - \eta$. Using (3.2) we can choose $\eta > 0$ sufficienty large and $\delta > 0$ so that $B_\delta(\eta_j) \subset A_\infty$ for all all $j = 1, \ldots, k$, and $B_\delta(\eta_j) \cap B_\varepsilon(v_\ell) = \varnothing$ for all $j \neq \ell$, as well as $|\rho(z) + \widetilde{r}(z) - \overline{z}| < |g(z)|$ for all $z \in \partial B_\delta(\eta_j)$, $j = 1, \ldots, k$. Therefore

$$|f_\eta(z) - g(z)| = |\rho(z) + \widetilde{r}(z) - \overline{z}| < |g(z)| \leq |f_\eta(z)| + |g(z)|$$

for all $z \in \partial B_\delta(\eta_j)$, and the application of Theorem 2.6 and Proposition 2.4 finishes the proof. □

At the end of this section we will show that the assertions of Theorem 3.1 hold for every $\eta \in \mathbb{C}$ with $|\eta|$ large enough.

We next prove that a *sufficiently small* shift $\eta \in \mathbb{C}$ changes neither the number nor the orientation of the regular zeros of f. This result is a slight extension of [11, Lemma 2.5].

Theorem 3.2. *Let f be as in* (3.1)*, and let $z_1, \ldots, z_m$ be the regular, and $z_{m+1}, \ldots, z_M$ be the singular zeros of f. Let further $\varepsilon > 0$ be such that $\left(\bigcup_{j=1}^m B_\varepsilon(z_j)\right) \cap \mathcal{C} = \varnothing$, and $B_\varepsilon(z_j) \cap B_\varepsilon(z_k) = \varnothing$ for all $j, k \in \{1, \ldots, M\}$ with $j \neq k$. If $\eta \in \mathbb{C}$ satisfies*

$$|\eta| < \delta := \min\Big\{|f(z)| : z \in \mathbb{C} \setminus \bigcup_{j=1}^M B_\varepsilon(z_j)\Big\},$$

then the following properties hold:

(i) For each $j = 1, \ldots, m$ the functions f and f_η have the same orientation on $B_\varepsilon(z_j)$, and $N(f; B_\varepsilon(z_j)) = N(f_\eta; B_\varepsilon(z_j)) = 1$.

(ii) $N(f; \mathbb{C} \setminus \bigcup_{j=m+1}^M B_\varepsilon(z_j)) = N(f_\eta; \mathbb{C} \setminus \bigcup_{j=m+1}^M B_\varepsilon(z_j)) = m$.

Proof. (i) From the construction it is clear that f and f_η have the same orientation on each set $B_\varepsilon(z_j)$. Moreover, for all $z \in \partial B_\varepsilon(z_j)$ we have

$$|f(z) - f_\eta(z)| = |\eta| < \delta \le |f(z)| \le |f(z)| + |f_\eta(z)|, \tag{3.3}$$

and hence $V(f; \partial B_\varepsilon(z_j)) = V(f_\eta; \partial B_\varepsilon(z_j))$ by Theorem 2.6. Since f has exactly one zero in $B_\varepsilon(z_j)$, and the poles of f and f_η coincide, the assertion follows from Corollary 2.5.

(ii) We know from (i) that f_η has exactly one zero in in each of the sets $B_\varepsilon(z_j)$, $j = 1, \dots, m$. If f_η has an additional zero $z_0 \in \mathbb{C} \setminus \bigcup_{j=m+1}^{M} B_\varepsilon(z_j)$, then we can choose an $\varepsilon_0 > 0$ such that f is nonzero on $\overline{B_{\varepsilon_0}(z_0)}$. Then (3.3) holds for all $z \in \partial B_{\varepsilon_0}(z_0)$, and Theorem 2.6 yields $V(f; \partial B_{\varepsilon_0}(z_0)) = V(f_\eta; \partial B_{\varepsilon_0}(z_0))$, which is a contradiction, since f and f_η have the same number of poles, but f has no zeros in $B_{\varepsilon_0}(z_0)$. □

Note that our general assumption $\lim_{|z|\to\infty} |f(z)| = \infty$ implies that $\delta > 0$ in Theorem 3.2.

Our next goal is to show that the number of zeros of the shifted functions f_η remains constant as long as the shift η does not cross a caustic of f. Our proof is based on the following two lemmas.

Lemma 3.3. *If Γ is a critical curve of f, and $\eta_1, \eta_2 \in \mathbb{C}$ are such that $\lambda\eta_1 + (1-\lambda)\eta_2 \notin f(\Gamma)$ holds for all $0 \le \lambda \le 1$, then $V(f_{\eta_1}; \Gamma) = V(f_{\eta_2}; \Gamma)$.*

Proof. Using an appropriate rotation and translation of the complex plane we may assume without loss of generality that $\eta_1 > 0$ and $\eta_2 = 0$. Our assumption then reads $\lambda\eta_1 \notin f(\Gamma)$ for all $0 \le \lambda \le 1$, and Proposition 2.2 implies that $f_{\lambda\eta_1}(z) = f(z) - \lambda\eta_1 \neq 0$ holds for all $0 \le \lambda \le 1$ and $z \in \Gamma$.

By construction and the triangle inequality we have

$$\eta_1 = |f(z) - f_{\eta_1}(z)| \le |f(z)| + |f_{\eta_1}(z)| \quad \text{for all } z \in \mathbb{C}. \tag{3.4}$$

If equality holds in (3.4) for some $z_0 \in \Gamma$, then $|f(z_0)| < \eta_1$, since $f_{\eta_1}(z_0) \neq 0$. Moreover,

$$\eta_1 - |f(z_0)| = |f_{\eta_1}(z_0)| = |\eta_1 - f(z_0)|,$$

which implies, together with $|f(z_0)| < \eta_1$, that $f(z_0) = \mu\eta_1$ for some $0 < \mu < 1$. But this means that $f(z_0) - \mu\eta_1 = 0$ with $z_0 \in \Gamma$, i.e., $\mu\eta_1$ is on the caustic $f(\Gamma)$, which is a contradiction. Consequently, we must have a strict inequality in (3.4), and hence $V(f_{\eta_1}; \Gamma) = V(f; \Gamma)$ by Theorem 2.6. □

Lemma 3.4. *If $\eta_1, \eta_2 \in \mathbb{C}$ are such that $\lambda\eta_1 + (1-\lambda)\eta_2 \notin f(\mathcal{C})$ holds for all $0 \le \lambda \le 1$, then $N(f_{\eta_1}; A) = N(f_{\eta_2}; A)$ holds for each set $A \in \mathcal{A}$, and $N(f_{\eta_1}) = N(f_{\eta_2})$.*

Proof. By Proposition 2.2, the functions f_{η_1} and f_{η_2} are regular. Moreover, these functions have the same poles, which are equal to of poles of f. For a bounded set $A \in \mathcal{A}$ we have a unique critical curve $\Gamma^* \subset \partial A$ such that $A \subset \operatorname{int}(\Gamma^*)$. If f is sense-preserving on A, then Corollary 2.5 and Lemma 3.3 imply

$$\begin{aligned} N(f_{\eta_1}; A) &= V(f_{\eta_1}; \Gamma^*) - \sum_{\substack{\Gamma \subset \partial A \text{ crit. curve,} \\ \operatorname{int}(\Gamma) \cap A = \varnothing}} V(f_{\eta_1}; \Gamma) - P(f; \operatorname{int}(\Gamma)) \\ &= V(f_{\eta_2}; \Gamma^*) - \sum_{\substack{\Gamma \subset \partial A \text{ crit. curve,} \\ \operatorname{int}(\Gamma) \cap A = \varnothing}} V(f_{\eta_2}; \Gamma) - P(f; \operatorname{int}(\Gamma)) = N(f_{\eta_2}; A). \end{aligned}$$

If f is sense-reversing on A, then

$$\begin{aligned} N(f_{\eta_1}; A) &= -V(f_{\eta_1}; \Gamma^*) + \sum_{\substack{\Gamma \subset \partial A \text{ crit. curve,} \\ \operatorname{int}(\Gamma) \cap A = \varnothing}} V(f_{\eta_1}; \Gamma) \\ &= -V(f_{\eta_2}; \Gamma^*) + \sum_{\substack{\Gamma \subset \partial A \text{ crit. curve,} \\ \operatorname{int}(\Gamma) \cap A = \varnothing}} V(f_{\eta_2}; \Gamma) = N(f_{\eta_2}; A). \end{aligned}$$

Using an additional artificial curve $\partial B_R(0)$ for a sufficiently large $R > 0$, containing all zeros and poles of f_{η_1} and f_{η_2} in its interior, we obtain the equality for the set A_∞. Finally, $N(f_{\eta_1}) = N(f_{\eta_2})$ holds since f_{η_1} and f_{η_2} have no singular zeros; cf. Proposition 2.2. □

Now suppose that $\eta_1, \eta_2 \in \mathbb{C}$ are linked by a continuous path, $\psi : [0,1] \to \mathbb{C}$, with $\psi(0) = \eta_1$, $\psi(1) = \eta_2$, and $\psi([0,1]) \cap f(\mathcal{C}) = \varnothing$. Since $\mathbb{C} \setminus f(\mathcal{C})$ is open, we can approximate $\psi([0,1])$ arbitrarily closely by a polygonal chain in $\mathbb{C} \setminus f(\mathcal{C})$; see Figure 2 for an illustration. Applying Lemma 3.4 successively on this chain gives the following result.

Theorem 3.5. *If $\eta_1, \eta_2 \in \mathbb{C}$ are linked by continuous path that does not cross a caustic of f, then $N(f_{\eta_1}; A) = N(f_{\eta_2}; A)$ holds for each set $A \in \mathcal{A}$, and $N(f_{\eta_1}) = N(f_{\eta_2})$.*

Figure 3 illustrates Theorem 3.5. In the left and middle plot we see the zeros, poles and critical curves of a function f for two shifts η_1 and η_2. Since there is a continuous path from η_1 to η_2, which does not cross a caustic of f (see the plot on the right), f_{η_1} and f_{η_2} have the same number of zeros, and these have the same locations with respect to the critical curves.

Moreover, Theorem 3.5 implies that in Theorem 3.1 we can replace $\eta > 0$ by any sufficiently large $\widetilde{\eta} \in \mathbb{C}$.

Remark 3.6. A rational harmoinc function f as in (3.1) is called *extremal*, when it has the maximum number of $5\deg(r) - 5$ zeros. As mentioned in

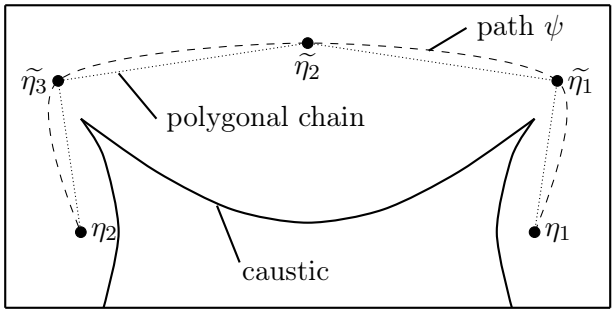


Figure 2: Illustration of Theorem 3.5.

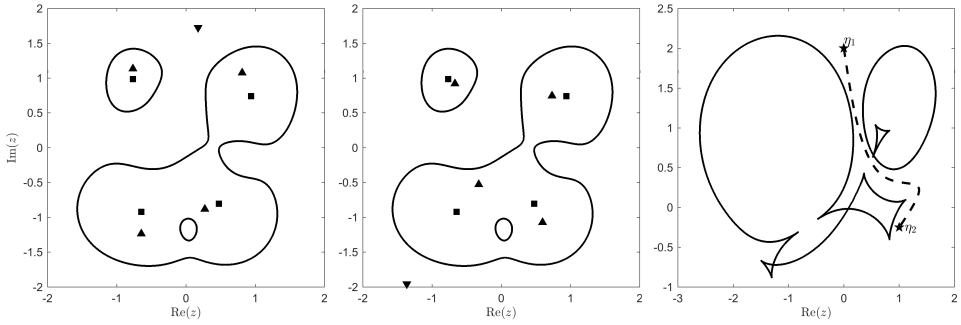


Figure 3: Left and middle: Critical curves (solid), sense-preserving zeros (▲), sense-reversing zeros (▼), and poles (■) for f_{η_1} (left) and f_{η_2} (middle). Right: Caustics and the corresponding shift (dashed).

the Introduction, an explicit construction of Rhie [17] yields an extremal function f with r as in (1.2) and $\eta = 0$ for each $n = \deg(r) \geq 2$. We then have $r = p/q$, where $\deg(p) = n - 1$ and $\deg(q) = n$. Our Theorems 3.1 and 3.5 imply that whenever $|\eta|$ is large enough, the shifted function f_η has exactly $n + 1$ zeros, namely n zeros close to the poles of f, and one zero in A_∞. Thus, f_η has $4n - 6$ fewer zeros than the extremal function f. An example of an extremal rational harmonic function f with $\deg(r) = 3$ and hence 10 zeros is shown in Figure 8. In that example a sufficiently large $|\eta|$ leads to a function f_η with only 4 zeros.

4 Crossing a caustic of f

In this section we will investigate the situation when a constant shift results in a caustic crossing of a function f as in (3.1). Let $z_0 \in \mathcal{C}$ be a critical point of f, i.e., $|r'(z_0)| = 1$, and let us define $\eta := f(z_0)$, so that z_0 is a singular zero of $f_\eta(z) = f(z) - \eta$. Using the Taylor series of $r(z)$ at $z = z_0$

and $r(z_0) - \eta = \overline{z}_0$, we then have

$$f_\eta(z) = r'(z_0)(z-z_0) + \frac{r''(z_0)}{2}(z-z_0)^2 - (\overline{z}-\overline{z}_0) + \sum_{k=3}^{\infty} \frac{r^{(k)}(z_0)}{k!}(z-z_0)^k. \quad (4.1)$$

For simplicity of notation we will now assume that

$$z_0 = 0 \quad \text{and} \quad r'(0) = 1.$$

This assumption amounts to a shift and rotation of the complex plane and hence it can be made without loss of generality of the results on the zeros of f_η that we will derive in the following. Under our assumption we can write (4.1) as

$$f_\eta(z) = T(z) + R(z), \quad \text{where} \quad (4.2)$$

$$T(z) := dz^2 + z - \overline{z}, \quad d := \frac{r''(0)}{2}, \quad R(z) := \sum_{k=3}^{\infty} \frac{r^{(k)}(0)}{k!} z^k. \quad (4.3)$$

Because of the non-degeneracy assumption on f we have $r''(0) \neq 0$, and thus $d \neq 0$.

Our strategy in the following is to show that in the neighborhood of $z = 0$ the remainder term R is "small enough", so that the zeros of f_η are close to the zeros of T, which can be explicitly analyzed. This approach is similar in spirit to the perturbation analysis in [21]. Note that since T is a harmonic polynomial of degree 2, it has at most 4 zeros [7].

Lemma 4.1. *For a given $\delta \in \mathbb{R}$, let $T_{\delta d}(z) := T(z) - \delta d$. Then all real zeros of $T_{\delta d}$ are given by*

$$z = \begin{cases} 0, & \text{if } \delta = 0, \\ \pm\sqrt{\delta}, & \text{if } \delta > 0, \end{cases} \quad (4.4)$$

and all non-real zeros of $T_{\delta d}$ are given by

$$z = \begin{cases} -d^{-1} \pm i\sqrt{|d^{-2}| - \delta}, & \text{if } |d^{-2}| \geq \delta \text{ and } \operatorname{Re}(d^{-1})^2 \neq \delta, \\ -d^{-1} - i\operatorname{Im}(d^{-1}), & \text{if } |d^{-2}| \geq \delta,\ \operatorname{Re}(d^{-1})^2 = \delta \text{ and } \operatorname{Im}(d^{-1}) \neq 0. \end{cases} \quad (4.5)$$

In particular, if z is a non-real zero of $T_{\delta d}(z)$, then $|z| \geq |\operatorname{Re}(d^{-1})| > 0$.

Proof. Let us write $z = x + iy$ and $d^{-1} = \alpha + i\beta$. The equation $T_{\delta d}(z) = 0$ holds if and only if

$$z^2 + d^{-1}(z - \overline{z}) - \delta = 0.$$

Splitting this equation into its real and imaginary parts gives the two equations

$$0 = y^2 + 2\beta y - x^2 + \delta, \quad (4.6)$$

$$0 = xy + \alpha y, \quad (4.7)$$

which we need to solve for real x and y.

If $y = 0$, then (4.6) implies $x^2 = \delta$, where both x and δ are real. Thus, all solutions of $T_{\delta d}(z) = 0$ with $\operatorname{Im}(z) = 0$ are given by $z = 0$ if $\delta = 0$, and $z = \pm\sqrt{\delta}$ if $\delta > 0$. If $\delta < 0$, then there exists no real solution.

If $y \neq 0$, then (4.7) implies $x = -\alpha$, and substituting $x^2 = \alpha^2$ in (4.6) yields

$$y_\pm = -\beta \pm \sqrt{\beta^2 + \alpha^2 - \delta}.$$

A solution of $T_{\delta d}(z) = 0$ with $\operatorname{Im}(z) \neq 0$ exists only when $y_\pm \in \mathbb{R} \setminus \{0\}$.

We have $y_\pm \in \mathbb{R}$ if and only if $\beta^2 + \alpha^2 - \delta = |d^{-2}| - \delta \geq 0$. If this holds, and we additionally have $\alpha^2 - \delta = \operatorname{Re}(d^{-1})^2 - \delta \neq 0$, then $y_\pm \neq 0$, and $T_{\delta d}(z) = 0$ has the two non-real solutions

$$z_\pm = -\alpha + iy_\pm = -d^{-1} \pm i\sqrt{|d^{-2}| - \delta}.$$

If $\alpha^2 - \delta = 0$ and $\beta \neq 0$, then $y_+ = 0$ and $y_- = -2\beta \neq 0$, so that

$$z_- = -\alpha + iy_- = -d^{-1} - i\operatorname{Im}(d^{-1})$$

is the only non-real solution. If $\alpha^2 - \delta = 0$ and $\beta = 0$, then there exists no non-real solution. □

Remark 4.2. Lemma 4.1 gives a complete characterization of all choices of $\delta \in \mathbb{R}$ that lead to an *extremal* harmonic polynomial $T_{\delta d}$ that has the maximum number of 4 zeros. For such a polynomial we need $0 < \delta \leq |d^{-2}|$ and $\delta \neq \operatorname{Re}(d^{-1})^2$, and then the 4 zeros are $z = \pm\sqrt{\delta}$ and $z = -d^{-1} \pm i\sqrt{|d^{-1}| - \delta}$.

Each non-real zero of $T_{\delta d}$ satisfies $|z| \geq |\operatorname{Re}(d^{-1})| > 0$, independently of the size of δ. Thus, if $\operatorname{Re}(d) \neq 0$, then for $\delta > 0$ small enough, the only zeros of $T_{\delta d}$ in a (small enough) neighborhood of $z = 0$ are the two real zeros $\pm\sqrt{\delta}$. This fact will be very important in the proof of the following result.

Theorem 4.3. *Let f be as in (3.1) with $r'(0) = 1$, suppose that the fold point $\eta := f(0)$ is simple, and let $A_+, A_- \in \mathcal{A}$ be the bordered sets on the critical point $z = 0$. Then there exists a nonzero $\widetilde{\eta} \in \mathbb{C}$, such that for all $0 < \alpha \leq 1$ we have*

(i) $N(f_{\eta+\alpha\widetilde{\eta}}) = N(f_\eta) + 1 = N(f_{\eta-\alpha\widetilde{\eta}}) + 2$,
(ii) $N(f_{\eta+\alpha\widetilde{\eta}}; A) = N(f_\eta; A) = N(f_{\eta-\alpha\widetilde{\eta}}; A)$ *for all* $A \in \mathcal{A} \setminus \{A_+, A_-\}$,
(iii) $N(f_{\eta+\alpha\widetilde{\eta}}; A_+) = N(f_{\eta-\alpha\widetilde{\eta}}; A_+) + 1$,
(iv) $N(f_{\eta+\alpha\widetilde{\eta}}; A_-) = N(f_{\eta-\alpha\widetilde{\eta}}; A_-) + 1$,
(v) $N_s(f_{\eta+\alpha\widetilde{\eta}}) = N_s(f_{\eta-\alpha\widetilde{\eta}}) = 0$, *and* $N_s(f_\eta) = 1$.

Proof. We will write f_η as in (4.2)–(4.3), and for a given $\delta > 0$ we will write $T_{\pm\delta d}(z) := T(z) \mp \delta d$. Since $f(0)$ is a (simple) fold point, we have $\operatorname{Re}(d) \neq 0$; see Lemma 2.1.

We know that if $\widetilde{\delta} > 0$ is small enough, then there exists an $\widetilde{\varepsilon}$, depending on $\widetilde{\delta}$ and with $\widetilde{\varepsilon} > \sqrt{\widetilde{\delta}} > 0$, such that the only zeros of $T_{+\widetilde{\delta}d}$ in the open disk $B_{\widetilde{\varepsilon}}(0)$ are the two real zeros $z = \pm\sqrt{\widetilde{\delta}}$. The function $T_{-\widetilde{\delta}d}$ has no zeros in that disk. Moreover, by shrinking $\widetilde{\varepsilon}$ and $\widetilde{\delta}$ if necessary, we can assume that $f_{\eta\pm\widetilde{\delta}d}$ has no pole in $B_{\widetilde{\varepsilon}}(0)$, since $|f_{\eta\pm\widetilde{\delta}d}(0)| = |\widetilde{\delta}d| < \infty$.

The orientation of $T_{+\widetilde{\delta}d}$ is determined by its Jacobian

$$J_{T_{+\widetilde{\delta}d}}(z) = |2dz + 1|^2 - 1.$$

Thus, by possibly shrinking $\widetilde{\delta} > 0$ once more, we can assume that $T_{+\widetilde{\delta}d}$ is differently oriented at its two (real) zeros $z = \pm\sqrt{\widetilde{\delta}}$.

The main idea now is to suitably choose ε and δ with $\varepsilon > \sqrt{\delta} > 0$, by possibly further shrinking the values $\widetilde{\varepsilon}$ and $\widetilde{\delta}$ obtained above, so that we can successfully apply Theorem 2.6 to $f_{\eta\pm\delta d}$ and $T_{\pm\delta d}$ on the closed Jordan curves

$$\Gamma^+ := \left(B^+, B_\varepsilon(0) \cap \mathcal{C}\right) \quad \text{and} \quad \Gamma^- := \left(B^-, B_\varepsilon(0) \cap \mathcal{C}\right), \tag{4.8}$$

where $B^+ := \partial B_\varepsilon(0) \cap \Omega_+$ and $B^- := \partial B_\varepsilon(0) \cap \Omega_-$. Thus, we have to verify that

$$|f_{\eta\pm\delta d}(z) - T_{\pm\delta d}(z)| = |R(z)| < |f_{\eta\pm\delta d}(z)| + |T_{\pm\delta d}(z)| \tag{4.9}$$

for all $z \in \partial B_\varepsilon(0) \cup (B_\varepsilon(0) \cap \mathcal{C})$.

The following argument is quite technical since the constant shift $\pm\delta d$ and the radius ε influence each other.

In the neighborhood of $z = 0$ we have

$$|T(z)| \in \mathcal{O}(|z|) \quad \text{and} \quad |R(z)| \in \mathcal{O}(|z|^3),$$

and consequently

$$|R(z)| < |T(z)| \quad \text{in } B_{\varepsilon^*}(0) \text{ for a sufficiently small } \varepsilon^* > 0.$$

We can assume that $0 < \varepsilon^* \leq \widetilde{\varepsilon}$, and we now have to find a corresponding $\delta^* > 0$. To this end we define

$$\varepsilon(\delta) := \max\{0 \leq \varepsilon \leq \varepsilon^* : |R(z)| < |T(z)| - \delta|d| \text{ for all } z \in \partial B_\varepsilon(0)\},$$

which is a continuous function of the real variable $\delta \geq 0$. We have $\varepsilon(0) = \varepsilon^* > 0$ and $\lim_{\delta\to\infty} \varepsilon(\delta) = 0$, where $B_0(0) = \varnothing$. Thus, for every continuous and strictly monotonically increasing function

$$\psi : [0, \infty) \to [0, \infty) \quad \text{with} \quad \psi(0) = 0,$$

there exists a $\delta^* > 0$, such that $\varepsilon(\delta^*) > \psi(\delta^*)$. Using the function $\psi(t) := \sqrt{t}$ yields parameters δ^* and $\varepsilon(\delta^*)$ with $\widetilde{\varepsilon} > \varepsilon(\delta^*) > \sqrt{\delta^*} > 0$, so that only the two real zeros $z_\pm = \pm\sqrt{\delta^*}$ of $T_{+\delta^* d}$ lie in the disk $B_{\varepsilon(\delta^*)}(0)$, and $T_{+\delta^* d}$ is differently oriented at these zeros.

For all $z \in \partial B_{\varepsilon(\delta^*)}(0)$ we immediately obtain

$$|f_{\eta\pm\delta^* d}(z) - T_{\pm\delta^* d}(z)| = |R(z)| < |T(z)| - \delta^*|d| \leq |f_{\eta\pm\delta^* d}(z)| + |T_{\pm\delta^* d}(z)|.$$

We also have to verify inequality (4.9) on $B_{\varepsilon(\delta^*)}(0) \cap \mathcal{C}$. Using (2.2) with our assumptions $z_0 = 0$, $r'(0) = 1$, and $d = r''(0)/2$, we see that this curve is given by

$$g(t) = ht + \mathcal{O}(|t|^2), \quad h := \frac{i\overline{d}}{|d|}.$$

A straightforward computation shows that

$$T(ht) = -\overline{d}t^2 + 2i\frac{\operatorname{Re}(d)}{|d|}t.$$

For $\varepsilon^* > 0$ sufficiently small and $|t| \leq \sqrt{\delta^*/2}$ we obtain, by restricting to the real part,

$$\begin{aligned} |T_{\pm\delta^* d}(ht)| &= \left|-\overline{d}t^2 + 2i\frac{\operatorname{Re}(d)}{|d|}t \mp \delta^* d\right| \geq |\operatorname{Re}(d)| \cdot |t^2 \pm \delta^*| \\ &\geq \frac{|\operatorname{Re}(d)|\delta^*}{2} > c_1|t|^3 \geq |R(ht)|\,, \end{aligned}$$

where $c_1 > 0$ is a real constant, and we have used that $\operatorname{Re}(d) \neq 0$.

On the other hand, for $\varepsilon^* > 0$ sufficiently small and $\sqrt{\delta^*/2} \leq |t| \leq \varepsilon(\delta^*)$ we obtain, by restricting to the imaginary part,

$$\begin{aligned} |T_{\pm\delta^* d}(ht)| &= \left|-\overline{d}t^2 + 2i\frac{\operatorname{Re}(d)}{|d|}t \mp \delta^* d\right| \geq \left|\operatorname{Im}(d)(t^2 \mp \delta^*) + 2\frac{\operatorname{Re}(d)}{|d|}t\right| \\ &\geq c_2\sqrt{\frac{\delta^*}{2}} > c_3|t|^3 \geq |R(ht)|\,, \end{aligned}$$

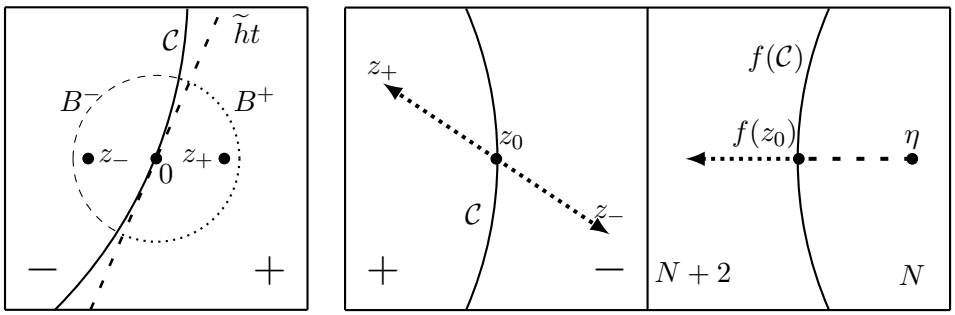
where $c_2, c_3 > 0$ are real constants. Note that in order to obtain the second inequality, it is again necessary that $\operatorname{Re}(d) \neq 0$.

Together we have

$$|T_{\pm\delta^* d}(ht)| > |R(ht)| \quad \text{for} \quad |t| < \varepsilon(\delta^*). \tag{4.10}$$

Clearly, if we do the same computations with $g(t) = ht + \mathcal{O}(|t|^2)$ instead of ht, we obtain the same estimate as in (4.10) for a possibly smaller $\varepsilon^* > 0$. Hence,

$$\begin{aligned} |f_{\eta\pm\delta^* d}(g(t)) - T_{\pm\delta^* d}(g(t))| = |R(g(t))| &< |T_{\pm\delta^* d}(g(t))| \\ &\leq |T_{\pm\delta^* d}(g(t))| + |f_{\eta\pm\delta^* d}(g(t))| \end{aligned}$$



(a) Construction in the proof of Theorem 4.3.

(b) Crossing a caustic at a fold point (cf. [14, Figures 9.2 and 9.3]).

Figure 4: Local behavior near fold points.

holds for all t with $|g(t)| \leq \varepsilon(\delta^*)$. Consequently, (4.9) is fulfilled for all $z \in B_{\varepsilon(\delta^*)}(0) \cap \mathcal{C}$.

In summary, we can apply Theorem 2.6 on Γ^+ and Γ^-, see (4.8). With Corollary 2.5 this yields

$$\begin{aligned}
N(f_{\eta+\delta^* d}; \operatorname{int}(\Gamma^+)) &= V(f_{\eta+\delta^* d}; \Gamma^+) = V(T_{+\delta^* d}; \Gamma^+) = 1,\\
N(f_{\eta+\delta^* d}; \operatorname{int}(\Gamma^-)) &= -V(f_{\eta+\delta^* d}; \Gamma^-) = -V(T_{+\delta^* d}; \Gamma^-) = 1,\\
N(f_{\eta-\delta^* d}; \operatorname{int}(\Gamma^+)) &= V(f_{\eta-\delta^* d}; \Gamma^+) = V(T_{-\delta^* d}; \Gamma^+) = 0,\\
N(f_{\eta-\delta^* d}; \operatorname{int}(\Gamma^-)) &= -V(f_{\eta-\delta^* d}; \Gamma^-) = -V(T_{-\delta^* d}; \Gamma^-) = 0.
\end{aligned}$$

Using Lemma 3.3 and Theorem 3.2 (again for possibly smaller $\varepsilon^* > 0$), we see that the assertions (iii) and (iv) are fulfilled for $\alpha = 1$ and $\widetilde{\eta} := \delta^* d$. The same argument as in the proof of Lemma 3.4 gives assertion (ii), and therefore also (i) and (v) follow (all for $\alpha = 1$).

Finally, the assertions (i)–(v) hold for all $0 < \alpha \leq 1$, since for sufficiently small $\delta > 0$ the line between $+\delta d$ and $-\delta d$ contains only a single caustic point of f. □

While we have formulated Theorem 4.3 for the critical point $z_0 = 0$ and for $r'(0) = 1$, it is clear that the result holds for any $z_0 \in \mathcal{C}$ and the corresponding value $r'(z_0) = e^{i\varphi}$, as long as $\eta = f(z_0)$ is a simple fold point. For a *multiple* fold point η, the set of corresponding critical points $f^{-1}(\eta)$ contains more than one element, and then the effect of Theorem 4.3 happens simultaneously at each of these critical points. An example can be seen in Figure 8, where one of the caustics has 9 double fold points. When the caustic is crossed at one of these points in a suitable direction, the number of zeros of the shifted functions changes by 4.

In the proof of Theorem 4.3, the crossing of the caustic at a (simple) fold point was done in the direction d, i.e., we considered a shift on the line from $\eta - \delta^* d$ to $\eta + \delta^* d$. Using Theorem 3.5, we easily see that crossing the caustic

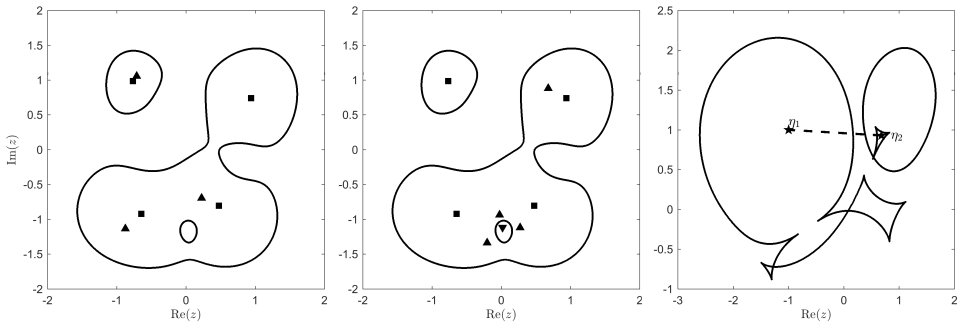


Figure 5: Left and middle: Critical curves (solid), sense-preserving zeros (▲), sense-reversing zeros (▼), and poles (■) for f_{η_1} (left) and f_{η_2} (middle). Right: Caustics and the corresponding shift (dashed).

in any other direction yields the same conclusion on the zeros of the shifted functions.

An illustration of the *local* behavior near a fold point is given in Figure 4(b). We shift the constant term η along the dotted line. Coming from the right, the function f_η has no zero close to the critical point z_0. For $\eta = f(z_0)$ there is exactly one (singular) zero of f_η, and after η crossed the caustic of f, a pair of differently oriented zeros of f_η appears.

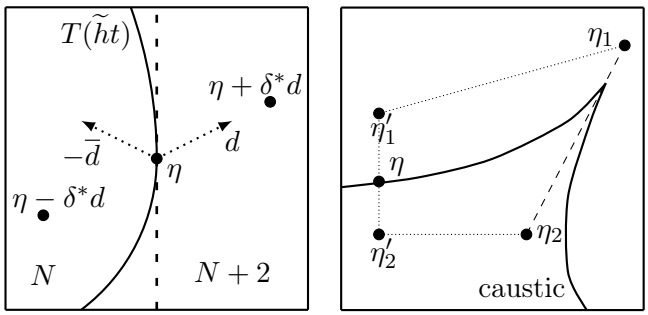
An illustration of the *global* effect of caustic crossings is shown in Figure 5. The plots on the left and in the middle show the critical curves, zeros, and poles of two functions f_{η_1} and f_{η_2}. On the right we plot the caustics and one possible path from η_1 to η_2. On every path from η_1 to η_2 we have at least three caustic crossings. With each crossing a pair of zeros in the neighborhood of the corresponding critical point appears or disappears. In this example we have a net gain of 2 zeros when traveling from η_1 to η_2, and a net loss of 2 zeros when traveling in the other direction.

The effect of 2 additional or 2 disappearing zeros is determined by the curvature of the caustic, which is given by the coefficient $-\overline{d}$ of the quadratic term of T, i.e., the caustic is locally a parabola. We have 2 additional zeros in case of crossing the caustic coming from the "open side" of the parabola, and 2 disappearing zeros coming from the other side; see Figure 5, Figure 6(b), and the examples in Section 5.

We are able to "simulate" the crossing of a cusp point using Theorem 3.5 and Theorem 4.3; see Figure 6(b). However, we also would like to give a local characterization of a cusp crossing. An important ingredient is the following result of Sheil-Small; see [23, Theorem 14].

Proposition 4.4. *If g is an analytic function in the convex domain D with $|g'(z)| < 1$ in D, then $g(z) - \overline{z}$ is univalent in D.*

If g is analytic and $|g'(z)| < 1$ in a star domain D with base point z_0,



(a) Relation between caustic curvature and the effect of constant shifts.

(b) Avoiding a cusp crossing.

Figure 6

then we can apply this proposition on the lines from z_0 to any point of D, which implies that $g(z) - \overline{z}$ attends the value $g(z_0) - \overline{z}_0$ exactly once in D. This fact will be used in the proof of the next theorem.

Theorem 4.5. *Let f be as in (3.1) with $r'(0) = 1$, suppose that the cusp point $\eta := f(0)$ is simple, and let $A_+, A_- \in \mathcal{A}$ be the bordered sets on the critical point $z = 0$. Then there exist a nonzero $\widetilde{\eta} \in \mathbb{C}$ and $b_+, b_- \in \{0, 1\}$ with $b_+ + b_- = 1$, such that for all $0 < \alpha \leq 1$ we have*

(i) $N(f_{\eta+\alpha\widetilde{\eta}}; A_+) = N(f_\eta; A_+) - b_+ = N(f_{\eta-\alpha\widetilde{\eta}}; A_+) + 1$,
(ii) $N(f_{\eta+\alpha\widetilde{\eta}}; A_-) = N(f_\eta; A_-) - b_- = N(f_{\eta-\alpha\widetilde{\eta}}; A_-) + 1$,
(iii) $N_s(f_{\eta+\alpha\widetilde{\eta}}) = N_s(f_{\eta-\alpha\widetilde{\eta}}) = 0$, *and* $N_s(f_\eta) = 1$.

Proof. The equalities

$$\begin{aligned} N(f_{\eta+\alpha\widetilde{\eta}}; A_+) &= N(f_{\eta-\alpha\widetilde{\eta}}; A_+) + 1, \\ N(f_{\eta+\alpha\widetilde{\eta}}; A_-) &= N(f_{\eta-\alpha\widetilde{\eta}}; A_-) + 1, \\ N_s(f_{\eta+\alpha\widetilde{\eta}}) &= N_s(f_{\eta-\alpha\widetilde{\eta}}) = 0, \end{aligned}$$

already follow from Theorem 4.3 and Lemma 3.4; see Figure 6(b). In order to show the remaining assertions we now investigate, as in the proof of Theorem 4.3, the functions $f_{\eta\pm\delta d}$ and $T_{\pm\delta d}$. Since we are in the cusp case, we have $\mathrm{Re}(d) = 0$ (see Lemma 2.1), and hence the non-real zeros of $T_{\pm\delta d}$ come into play.

From Lemma 4.1 we know that for all $0 < \delta < |d^{-2}|$, the function $T_{+\delta d}$ has the two real zeros $z = \pm\sqrt{\delta}$, while $T_{-\delta d}$ has no real zeros. Moreover, $T_{+\delta d}$ has the two purely imaginary zeros

$$z = -d^{-1} \pm i\sqrt{|d^{-2}| - \delta}, \tag{4.11}$$

and $T_{-\delta d}$ has the two purely imaginary zeros

$$z = -d^{-1} \pm i\sqrt{|d^{-2}| + \delta}. \tag{4.12}$$

Only one of the two zeros in (4.11) and in (4.12) is sufficiently close to $z = 0$, and the sign of $\mathrm{Im}(d^{-1})$ determines which one it is: If $\mathrm{Im}(d^{-1}) > 0$, then the zero of interest of $T_{\pm\delta d}$ is

$$z_\pm := -d^{-1} + i\sqrt{|d^{-2}| \mp \delta}$$

since then $|z_\pm| < \sqrt{\delta}$, while the other zero satisfies

$$|-d^{-1} - i\sqrt{|d^{-2}| \mp \delta}| \geq |d^{-1}| > \sqrt{\delta}.$$

From

$$\begin{aligned}|\partial_z T_{\pm\delta d}(z_\pm)| &= |2dz_\pm + 1| = |2d(-d^{-1} + i\sqrt{|d^{-2}| \mp \delta}) + 1| \\ &= |-1 + 2di\sqrt{|d^{-2}| \mp \delta}| = |-1 + 2\sqrt{1 \mp \delta|d|^2}|\end{aligned}$$

we see that z_+ is a sense-reversing zero of $T_{+\delta d}$, and z_- is a sense-preserving zero of $T_{-\delta d}$. For $\mathrm{Im}(d^{-1}) < 0$ we get an analogous result, but then the zeros in (4.11) and (4.12) change their roles, i.e., $z = -d^{-1} - i\sqrt{|d^{-2}| \mp \delta}$ is close to zero, and $z = -d^{-1} + i\sqrt{|d^{-2}| \mp \delta}$ is bounded away from zero.

We will now show that the zero $z_\pm$ of $T_{\pm\delta d}$ corresponds to a zero of $f_{\eta\pm\delta d}$ by applying Theorem 2.6 on $\partial B_{\widetilde{\varepsilon}}(z_\pm)$ for an appropriately chosen $\widetilde{\varepsilon} > 0$. For each $\varphi \in [0, 2\pi)$ we have

$$T_{\pm\delta d}(z_\pm + \widetilde{\varepsilon}e^{i\varphi}) = \widetilde{\varepsilon}\left(2dz_\pm e^{i\varphi} + d\widetilde{\varepsilon}e^{2i\varphi} + 2i\,\mathrm{Im}(e^{i\varphi})\right).$$

For a sufficiently small $\delta > 0$, which determines $z_\pm$, we now set $\widetilde{\varepsilon} := |z_\pm|$, and we assume that $2|d|\widetilde{\varepsilon} \leq 1$. Then

$$\begin{aligned}\left|T_{\pm\delta d}(z_\pm + \widetilde{\varepsilon}e^{i\varphi})\right| &\geq \widetilde{\varepsilon}\left(2\left|\pm|d|\widetilde{\varepsilon}e^{i\varphi} + i\,\mathrm{Im}(e^{i\varphi})\right| - |d|\widetilde{\varepsilon}\right) \\ &\geq \widetilde{\varepsilon}\left(2\left|\pm|d|\widetilde{\varepsilon}\,\mathrm{Re}(e^{i\varphi}) + (1 \pm |d|\widetilde{\varepsilon})i\,\mathrm{Im}(e^{i\varphi})\right| - |d|\widetilde{\varepsilon}\right) \\ &\geq \widetilde{\varepsilon}\left(2\left|\pm|d|\widetilde{\varepsilon}\,\mathrm{Re}(e^{i\varphi}) + |d|\widetilde{\varepsilon}i\,\mathrm{Im}(e^{i\varphi})\right| - |d|\widetilde{\varepsilon}\right) \\ &\geq \widetilde{\varepsilon}\,(2|d|\widetilde{\varepsilon} - |d|\widetilde{\varepsilon}) = |d|\widetilde{\varepsilon}^2 \\ &> c|z_{\mathrm{Im}} + \widetilde{\varepsilon}e^{i\varphi}|^3 \geq |R(z_{\mathrm{Im}} + \widetilde{\varepsilon}e^{i\varphi})|,\end{aligned}$$

for some constant $c > 0$. Thus, we have

$$|f_{\eta\pm\delta d}(z) - T_{\pm\delta d}(z)| = |R(z)| < |T_{\pm\delta d}(z)| + |f_{\eta\pm\delta d}(z)| \quad \text{for all } z \in \partial B_{\widetilde{\varepsilon}}(z_\pm).$$

Using Theorem 2.6 gives

$$V(f_{\eta\pm\delta d}; \partial B_{\widetilde{\varepsilon}}(z_\pm)) = V(T_{\pm\delta d}; \partial B_{\widetilde{\varepsilon}}(z_\pm)),$$

and, as a consequence,

$$N(f_{\eta\pm\delta d}; B_{\widetilde{\varepsilon}}(z_\pm)) = N(T_{\pm\delta d}; B_{\widetilde{\varepsilon}}(z_\pm)) = 1.$$

In the following we denote by $\widetilde{z}_\pm$ the zero of $f_{\eta\pm\delta d}$ corresponding to the zero $z_\pm$ of $T_{\pm\delta d}$. By construction, $\widetilde{z}_+$ is a sense-reversing zero of $f_{\eta+\delta^* d}$ and $\widetilde{z}_-$ is a sense-preserving zero of $f_{\eta-\delta^* d}$.

We now construct $\varepsilon, \delta > 0$ such that we can apply Theorem 2.6 on $\partial B_\varepsilon(0)$ and the zeros $\widetilde{z}_\pm$ of $f_{\eta\pm\delta d}$ are in $B_\varepsilon(0)$. Let $\varepsilon > 0$ be such that $f_\eta(z) \neq 0$ for all $z \in B_\varepsilon(0) \setminus \{0\}$, and f_η has no poles in $B_\varepsilon(0)$. Furthermore we define

$$\widetilde{\delta} := \min\left\{ \frac{|f_\eta(z)|}{|d|} : z \in \partial B_\varepsilon(0) \right\} \quad \text{and} \quad \delta := \frac{1}{2}\min\{\widetilde{\delta}, \varepsilon^2\}.$$

Hence, we have

$$|f_\eta(z) - f_{\eta\pm\delta d}(z)| = \delta|d| < |f_\eta(z)| + |f_{\eta\pm\delta d}(z)| \quad \text{for all } z \in \partial B_\varepsilon(0).$$

With Theorem 2.6 we get

$$V(f_{\eta\pm\delta d}; \partial B_\varepsilon(0)) = V(f_\eta; \partial B_\varepsilon(0)) \quad \text{and} \quad \widetilde{z}_\pm \in B_\varepsilon(0). \tag{4.13}$$

Now we look at the number of zeros of $f_{\eta\pm\delta d}$. Since $B_\varepsilon(0) \cap \Omega_-$ is a star domain with base point $\widetilde{z}_+$, the function $f_{\eta+\delta d}$ has no other zero than $\widetilde{z}_+$ in this domain; see Proposition 4.4 and its discussion. Consequently, the function $f_{\eta-\delta d}$, which results from crossing the caustic of f through the cusp point η, has either no ($b_- = 0$) or two ($b_- = 1$) zeros in $B_\varepsilon(0) \cap \Omega_-$; cf. Theorem 4.3. Furthermore, because of (4.13), the function $f_{\eta-\delta d}$ has either two ($b_+ = 1$) or no ($b_+ = 0$) fewer zeros than $f_{\eta+\delta d}$ in $B_\varepsilon(0) \cap \Omega_+$. Together this implies the remaining equalities in (i) and (ii) for $\alpha = 1$.

Finally, the assertions (i)–(v) hold for all $0 < \alpha \leq 1$, since for sufficiently small $\delta > 0$ the line between $+\delta d$ and $-\delta d$ contains only a single caustic point of f. □

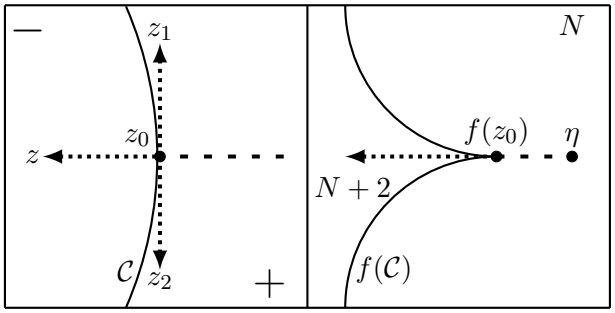
It is clear that Theorem 4.5 holds for an arbitrary $z_0 \in \mathcal{C}$, as long as $f(z_0)$ is a simple cusp point. For multiple points the effect happens again simultaneously at all corresponding critical points.

A cusp crossing is illustrated in Figure 7. We shortly describe the positive case. The constant term η is shifted along the dotted line. Coming from the right the function has only one sense-preserving zero close to z_0. When η reaches the caustic, the unique zero becomes singular. When η crosses the caustic, the initial zero crosses the critical curve and thus changes the orientation, i.e., it is now sense-reversing. Furthermore an additional pair of sense-preserving zeros appears. Hence we have three zeros after the caustic crossing. The same happens in the negative case with the reverse orientation.

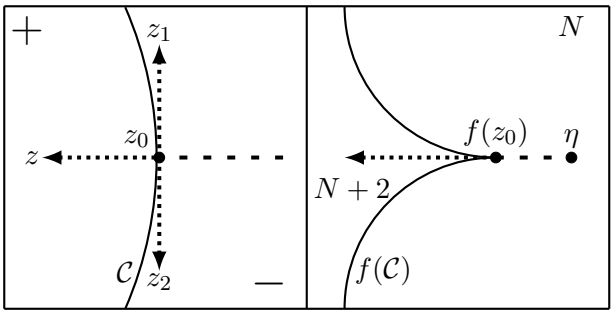
Finally, it is worth to point out that our results yield a characterization of the Poincaré index of a singular zero.

Corollary 4.6. *If f is as in* (3.1)*, and $z_0 \in \mathbb{C}$ is a singular zero of f, then*

$$\operatorname{ind}(f; z_0) = \begin{cases} 0, & \text{if } 0 \text{ is a fold point,} \\ \pm 1, & \text{if } 0 \text{ is a cusp point.} \end{cases}$$



(a) Crossing a positive cusp (cf. [14, Figure 9.7]).



(b) Crossing a negative cusp (cf. [14, Figure 9.8]).

Figure 7: Crossing a caustic at a cusp point.

Proof. Let $z_0 \in \mathbb{C}$ be a singular zero of f and choose some $\varepsilon > 0$, such that f has no other zero in $\overline{B_\varepsilon(z_0)}$. Using the same idea as in the proof of Theorem 3.2, we define

$$\delta := \min\{|f(z)| : z \in \partial B_\varepsilon(z_0)\}.$$

Then for each $\eta \in B_\delta(0)$ we have

$$|f(z) - f_\eta(z)| = |\eta| < \delta \leq |f(z)| \leq |f(z)| + |f_\eta(z)|$$

for all $z \in \partial B_\varepsilon(z_0)$, and Theorem 2.6 implies that

$$V(f; \partial B_\varepsilon(z_0)) = V(f_\eta; \partial B_\varepsilon(z_0)).$$

The assertion now follows from the proofs of Theorems 4.3 (fold case) and 4.5 (cusp case); see also the Figures 4(b) and 7. □

The cases $+1$ or -1, i.e., for positive or negative cusps, are determined by b_+ and b_- in Theorem 4.5. We see that the Poincaré index of a singular zero z_0 is the sum of the Poincaré indices of the regular zeros merging in z_0. Recently the Poincaré index of singular zeros of harmonic functions $h(z) - \overline{z}$, with a general analytic function h, were studied in [9] using the power series of h. However, a characterization whether the index is $+1$ or -1 in the cusp case is pointed out as future work.

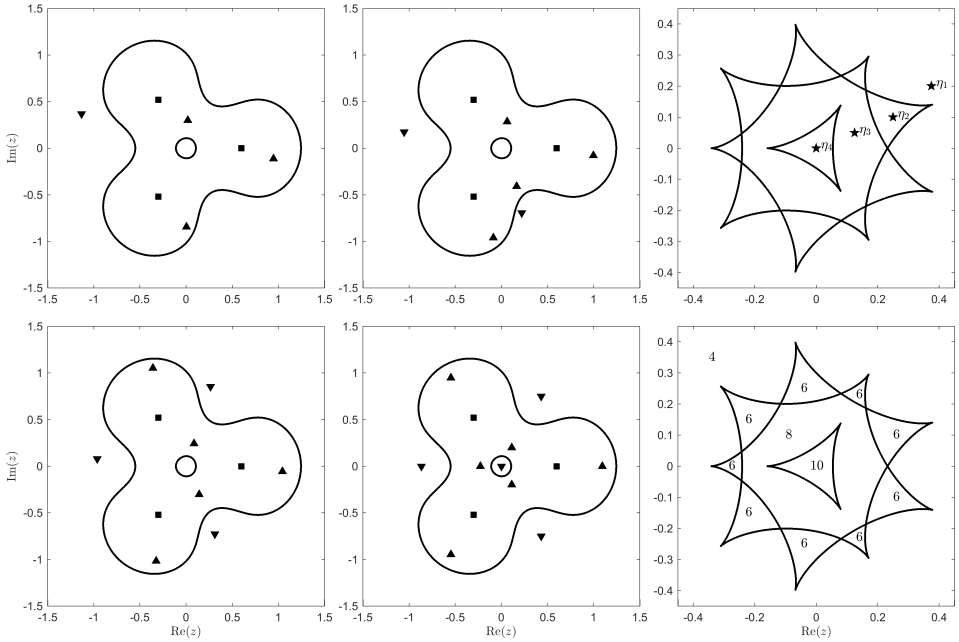


Figure 8: Critical curves, zeros (▲,▼) and poles (■) of f_{η_1} (top left), f_{η_2} (top mid), f_{η_3} (bottom left) and f_{η_4} (bottom mid); caustics (top right) and the number of zeros depending on the constant term η_j (bottom right).

5 Examples and outlook

Let us give some examples that illustrate the results of the previous sections.

First we consider the function

$$f(z) = \frac{z^{n-1}}{z^n - \rho^n} - \overline{z}$$

for some $\rho > 0$. Functions of this form have been frequently studied in the context of gravitational lensing; see, e.g., the original work of Mao, Petters and Witt [12], and the more recent articles [10, 20], which contain many further references. We choose $n = 3$ and $\rho = \frac{3}{5}$, and plot the zeros of f_η for several constant shifts η in Figure 8.

We know from Theorem 3.1, that for $|\eta| \gg 1$ the function f_η has 3 zeros close to its 3 poles, and one zero in the set A_∞. This can be observed for the shift η_1. The shift from η_1 to η_2 results in a caustic crossing with one additional pair of zeros (one sense-preserving and one sense-reversing) appearing at the outer critical curve, as predicted by Theorem 4.3 and the curvature of the caustic. The same happens when shifting from η_2 to η_3. Finally, the shift from η_3 to η_4 results in an additional pair of zeros at the inner critical curve. Note that $N(f_{\eta_4}) = 10 = 5n - 5$. Hence f_{η_4} is an

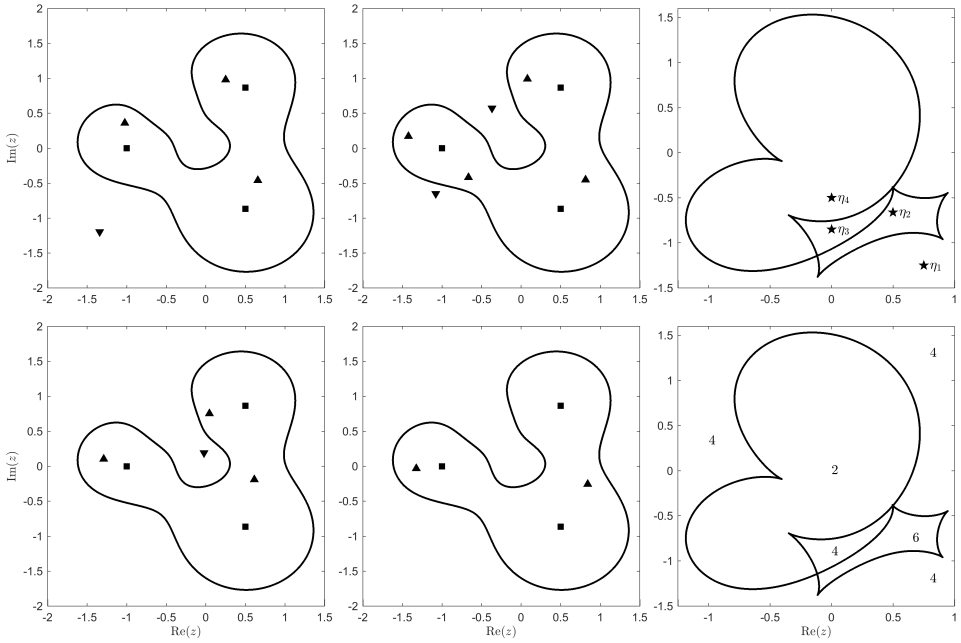


Figure 9: Critical curves, zeros (▲,▼) and poles (■) of f_{η_1} (top left), f_{η_2} (top mid), f_{η_3} (bottom left) and f_{η_4} (bottom mid); caustics (top right) and the number of zeros depending on the constant term η_j (bottom right).

extremal rational harmonic function, and it has $6 = 4n - 6$ more zeros than f_{η_1}; cf. Remark 3.6.

It was shown in [11, Theorem 3.1] (see also [8, Theorem 3.5]), that an extremal rational harmonic function is always regular, i.e., has no singular zeros. Our results in Section 4 yield the following slight generalization.

Lemma 5.1. *Let f be as in* (3.1)*, and suppose that there exists an $\varepsilon > 0$ with $N(f) \geq N(f_\eta)$ for all $\eta \in B_\varepsilon(0)$. Then f is regular.*

Proof. The function f is singular if and only if $z = 0$ is a caustic (fold or cusp) point of f. By the Theorems 4.3 (fold case) and 4.5 (cusp case) there exist some $\eta \in \mathbb{C}$ such that f_η has at least one additional zero, which contradicts the assumption $N(f) \geq N(f_\eta)$. □

Since an extremal rational harmonic function f satisfies $N(f) \geq N(f_\eta)$ for all $\eta \in \mathbb{C}$, Lemma 5.1 immediately implies that f must be regular. On the other hand, if f is singular, then for every $\varepsilon > 0$ there must exist an $\eta \in B_\varepsilon(0)$, such that f_η is regular and $N(f_\eta) \leq N(f)$.

As another example we consider

$$f(z) = \frac{(1+i)z^2 - i}{z^3 + 1} - \overline{z},$$

and plot the results in Figure 9. For η_1 we again have 3 zeros close to the 3 poles and one zero in A_∞, as shown by Theorem 3.1. The first caustic crossing from η_1 to η_2 results in one additional pair of zeros, but due to the curvature of the caustic, the shift from η_2 to η_3 reverses this effect. The last shift from η_3 to η_4 results again in two fewer zeros due to the curvature of the caustic, giving $N(f_{\eta_4}) = 2$. Since $N_-(f_{\eta_4}) = 0$, we have a rational harmonic function with the *minimal number of zeros*. (For $r = p/q$ with $\deg(p) \leq \deg(q)$ this number is $\deg(q) - 1$, which can be easily proved using the argument principle.)

Finally, we would like to mention that most of our theory in this paper can be extended from rational to general analytic functions, i.e., to functions of the form $f(z) = h(z) - \overline{z}$ with h being (locally) analytic. This is because the derivation of our main results is based on the local Taylor series, and in the more general case we we could start from

$$h(z) = h'(z_0)(z - z_0) + \frac{1}{2}h''(z_0)(z - z_0)^2 + R(z, z_0).$$

A similar approach has recently been used in [9].

Another interesting extension would be to consider rational harmonic functions of the form $r_1(z) - \overline{r_2(z)}$ with both r_1 and r_2 rational. We are not aware of any general results on the zeros of such functions.

Acknowledgements We thank Seung-Yeop Lee for sending us a pdf-file of [23].

References

[1] M. B. Balk, *Polyanalytic Functions*, vol. 63 of Mathematical Research, Akademie-Verlag, Berlin, 1991.

[2] K. Daněk and D. Heyrovský, *Image-plane analysis of n-point-mass lens critical curves and caustics*, The Astrophysical Journal, 806, Art. No. 63 (2015), pp. 1–14.

[3] P. Duren, W. Hengartner, and R. S. Laugesen, *The argument principle for harmonic functions*, Amer. Math. Monthly, 103 (1996), pp. 411–415.

[4] D. Khavinson, S.-Y. Lee, and A. Saez, *Zeros of harmonic polynomials, critical lemniscates, and caustics*, Complex Anal. Synerg., 4, Art. No. 2 (2018), pp. 1–20.

[5] D. Khavinson and G. Neumann, *On the number of zeros of certain rational harmonic functions*, Proc. Amer. Math. Soc., 134 (2006), pp. 1077–1085.

[6] D. Khavinson and G. Neumann, *From the fundamental theorem of algebra to astrophysics: a "harmonious" path*, Notices Amer. Math. Soc., 55 (2008), pp. 666–675.

[7] D. Khavinson and G. Świątek, *On the number of zeros of certain harmonic polynomials*, Proc. Amer. Math. Soc., 131 (2003), pp. 409–414.

[8] J. Liesen and J. Zur, *The maximal number of zeros of $r(z) - \overline{z}$ revisited*, Comput. Methods Funct. Theory, (2018). Online first publication.

[9] R. Luce and O. Sète, *The index of singular zeros of harmonic mappings of anti-analytic degree one*, Oberwolfach Report OWP 2017-03, (2017).

[10] R. Luce, O. Sète, and J. Liesen, *Sharp parameter bounds for certain maximal point lenses*, Gen. Relativity Gravitation, 46, Art. No. 1736 (2014), pp. 1–16.

[11] R. Luce, O. Sète, and J. Liesen, *A note on the maximum number of zeros of $r(z) - \overline{z}$*, Comput. Methods Funct. Theory, 15 (2015), pp. 439–448.

[12] S. Mao, A. O. Petters, and H. J. Witt, *Properties of point mass lenses on a regular polygon and the problem of maximum number of images*, in The Eighth Marcel Grossmann Meeting, Part A, B Jerusalem, 1997, World Sci. Publ., River Edge, NJ, 1999, pp. 1494–1496.

[13] A. O. Petters, *Gravity's action on light*, Notices Amer. Math. Soc., 57 (2010), pp. 1392–1409.

[14] A. O. Petters, H. Levine, and J. Wambsganss, *Singularity Theory and Gravitational Lensing*, vol. 21 of Progress in Mathematical Physics, Birkhäuser Boston, Inc., Boston, MA, 2001.

[15] A. O. Petters and M. C. Werner, *Mathematics of gravitational lensing: multiple imaging and magnification*, Gen. Relativity Gravitation, 42 (2010), pp. 2011–2046.

[16] A. O. Petters and H. J. Witt, *Bounds on number of cusps due to point mass gravitational lenses*, J. Math. Phys., 37 (1996), pp. 2920–2933.

[17] S. H. Rhie, *n-point gravitational lenses with 5(n-1) images*, arXiv:astro-ph/0305166v1, (2003).

[18] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses*, Springer-Verlag, Berlin, Heidelberg, 1992.

[19] P. Schneider and A. Weiss, *The two-point-mass lens - Detailed investigation of a special asymmetric gravitational lens*, Astronomy and Astrophysics, 164 (1986), pp. 237–259.

[20] O. Sète, R. Luce, and J. Liesen, *Creating images by adding masses to gravitational point lenses*, Gen. Relativity Gravitation, 47, Art. No. 42 (2015), pp. 1–8.

[21] O. Sète, R. Luce, and J. Liesen, *Perturbing rational harmonic functions by poles*, Comput. Methods Funct. Theory, 15 (2015), pp. 9–35.

[22] T. J. Suffridge and J. W. Thompson, *Local behavior of harmonic mappings*, Complex Variables Theory Appl., 41 (2000), pp. 63–80.

[23] A. S. Wilmshurst, *Complex Harmonic Mappings and the Valence of Harmonic Polynomials*, PhD thesis, Univ. of York, U.K., 1994.

[24] H. J. Witt and A. O. Petters, *Singularities of the one- and two-point mass gravitational lens*, J. Math. Phys., 34 (1993), pp. 4093–4111.

The maximum number of zeros of $r(z) - \overline{z}$ revisited*

Jörg Liesen† Jan Zur†

Abstract

Generalizing several previous results in the literature on rational harmonic functions, we derive bounds on the maximum number of zeros of functions $f(z) = \frac{p(z)}{q(z)} - \overline{z}$, which depend on both $\deg(p)$ and $\deg(q)$. Furthermore, we prove that any function that attains one of these upper bounds is regular.

Keywords: Zeros of rational harmonic functions, rational harmonic functions, harmonic polynomials, complex valued harmonic functions.

Mathematics Subject classification (2010): 30D05; 31A05; 37F10.

1 Introduction

We study the zeros of *rational harmonic functions* of the form

$$f(z) = r(z) - \overline{z} \quad \text{with} \quad r(z) = \frac{p(z)}{q(z)}, \tag{1.1}$$

where p and q are coprime polynomials of respective degrees n_p and n_q, and

$$n := \deg(r) = \max\{n_p, n_q\} \geq 2.$$

If $z_0 \in \mathbb{C}$ is a zero of f, i.e., $r(z_0) - \overline{z}_0 = 0$, then also $z_0 = \overline{r(z_0)}$. Inserting this into the first equation and taking complex conjugates gives $\overline{r}(r(z_0)) - z_0 = 0$. This can be transformed into a polynomial equation (of degree n^2+1), which shows that f has finitely many zeros. It is also important to note that because of the term $\overline{z}$, the zeros of f can not be "factored out", and hence they do

*A final version of this preprint manuscript appeared as follows: J. Liesen and J. Zur. The maximum number of zeros of $r(z) - \overline{z}$ revisited. *Comput. Methods Funct. Theory* (2018), 18(3), pp. 463–472, doi:10.1007/s40315-017-0231-1.

†TU Berlin, Institute of Mathematics, MA 4-5, Straße des 17. Juni 136, 10623 Berlin, Germany. {liesen,zur}@math.tu-berlin.de

not have a multiplicity in the usual sense. "Numbers of zeros" in this context therefore refer to numbers of distinct complex points.

Several authors have studied upper bounds on $N(f)$, the number of zeros of a function f as in (1.1). In particular, Khavinson and Neumann [4] showed that in general $N(f) \leq 5n - 5$, and Khavinson and Świątek [6] showed that if $n_q = 0$, i.e., f is a harmonic polynomial, then $N(f) \leq 3n - 2$. Results of Rhie [11] and Geyer [3], respectively, show that these two upper bounds are sharp for each $n \geq 2$.

The main purpose of this note is to prove the following theorem, which takes the individual degrees n_p and n_q into account, and which generalizes (almost) all previously known bounds on the maximal number of zeros of f.

Theorem 1.1. *Let f be as in* (1.1). *Then for every $c \in \mathbb{C}$, the number of zeros of $f_c(z) := f(z) - c$ satisfies*

$$N(f_c) \leq \begin{cases} 2n_p + 3n_q - 3, & \text{if } n_p < n_q, \\ 5n_p - 5, & \text{if } n_p = n_q, \\ 3n_p + 2n_q - 2, & \text{if } n_p > n_q + 1. \end{cases}$$

The only case "missing" in this theorem is $n_p = n_q + 1$. For this case we know from [9] that $N(f_c) \leq 5n_p - 6$.

Note that for each $c \in \mathbb{C}$ we can write

$$f_c(z) = r_c(z) - \overline{z}, \quad \text{where} \quad r_c(z) := \frac{p(z) - cq(z)}{q(z)}, \tag{1.2}$$

which is again a rational harmonic function of the form (1.1). The degree of the numerator polynomial of r_c is potentially different from n_p, and this allows for some flexibility in applications of Theorem 1.1. A second reason why we have formulated the result for the function f_c (rather than just f) is the application of rational harmonic functions in the context of gravitational lensing; see [5] for a survey. In that application a constant shift represents the position of the light source of the lens, and the change of the number of zeros under movements of the light source is of great interest; see, e.g., [8] for more details.

Our note is organized as follows. In Section 2 we briefly recall the mathematical background, in particular the argument principle for continuous functions and a helpful result from complex dynamics. In Section 3 we prove Theorem 1.1. We also prove that a function that attains the bound in Theorem 1.1 is regular, which generalizes a result from [10]. In Section 4 we explain why the special case $n_p = n_q + 1$ cannot be completely resolved by our method of proof, and we discuss the relation of Theorem 1.1 to all previously published bounds that we are aware of.

2 Mathematical background

Let f be as in (1.1). Using the Wirtinger derivatives ∂_z and $\partial_{\bar{z}}$ we can write the *Jacobian* of f as

$$J_f(z) = |\partial_z f(z)|^2 - |\partial_{\bar{z}} f(z)|^2 = |r'(z)|^2 - 1.$$

If $z_0 \in \mathbb{C}$ is a zero of f, i.e., $f(z_0) = 0$, then z_0 is called a *sense-preserving*, *sense-reversing*, or *singular zero* of f, if $|r'(z_0)| > 1$, $|r'(z_0)| < 1$, or $|r'(z_0)| = 1$, respectively. The sense-preserving and sense-reversing zeros of f are called the *regular zeros*. If f has only such zeros, then f is called *regular*, and otherwise f is called *singular*. We denote the number of sense-preserving, sense-reversing, and singular zeros of f in a set $S \subseteq \mathbb{C}$ by $N_+(f;S)$, $N_-(f;S)$, and $N_0(f;S)$, respectively. For $S = \mathbb{C}$ we simply write $N_+(f)$, $N_-(f)$, and $N_0(f)$. In our proofs we will use the following result on regular functions; see [4, Lemma].

Lemma 2.1. *If f is as in* (1.1)*, then the set of complex numbers c for which $f_c(z) = f(z) - c$ is regular, is open and dense in $\mathbb{C}$.*

This lemma can be easily shown when using the fact that the function f_c is singular if and only if c is a *caustic point* of f; see [8, Proposition 2.2].

Let Γ be a closed Jordan curve, and let g be any function that is continuous and nonzero on Γ. Then the *winding* of g on Γ is defined as the change in the argument of $g(z)$ as z travels once around Γ in the positive direction, divided by 2π, i.e.,

$$V(g;\Gamma) := \frac{1}{2\pi}\Delta_\Gamma \arg g(z).$$

The following result holds for the winding of the functions of our interest.

Theorem 2.2. *Let f be as in* (1.1)*. If f is nonzero and finite on a closed Jordan curve Γ and has no singular zero in* $\mathrm{int}(\Gamma)$*, then*

$$V(f;\Gamma) = N_+(f;\mathrm{int}(\Gamma)) - N_-(f;\mathrm{int}(\Gamma)) - P(f;\mathrm{int}(\Gamma)),$$

where $P(f;\mathrm{int}(\Gamma))$ denotes the number of poles with multiplicities of r, and hence of f, in $\mathrm{int}(\Gamma)$*.*

We will frequently use the following version of Rouché's theorem.

Theorem 2.3. *Let Γ be a closed Jordan curve and suppose that $f, g : \Gamma \to \mathbb{C}$ are continuous. If $|f(z) - g(z)| < |f(z)| + |g(z)|$ holds for all $z \in \Gamma$, then $V(f;\Gamma) = V(g;\Gamma)$.*

For more details on the mathematical background described above we refer to [4], [8] and [13].

In addition, we will need a result on fixed points from complex dynamics. Let $z_f \in \mathbb{C}$ be a fixed point of a rational function r, i.e., $r(z_f) = z_f$. Then z_f

is called *attracting*, *repelling*, or *rationally neutral*, if respectively $|r'(z_f)| < 1$, $|r'(z_f)| > 1$, or $|r'(z_f)| = 1$. The following is a combination of [2, Chapter III, Theorem 2.2 and 2.3].

Theorem 2.4. *If z_f is an attracting or rationally neutral fixed point of a rational function R with $\deg(R) \geq 2$, then exists a critical point z_c of R, i.e., $R'(z_c) = 0$, with $\lim_{k\to\infty} R^k(z_c) = z_f$, where $R^k := R \circ \cdots \circ R$ (k times).*

3 Main results

Our strategy to prove Theorem 1.1 is the following: First we determine $N_+(f_c) - N_-(f_c)$ for regular functions f as in (1.1) and $f_c(z) = f(z) - c$ with respect to n_p and n_q using Theorem 2.2 and 2.3. Then we bound $N_0(f) + N_-(f)$ for general f by the number of zeros of r' using Theorem 2.4. Finally, we combine both results with Lemma 2.1 in order to obtain the proof also for non-regular f.

We denote by $B_M(z_0)$ the open disk of radius $M > 0$ around $z_0 \in \mathbb{C}$.

Lemma 3.1. *Let f be as in (1.1) and suppose that f is regular, i.e., $N_0(f) = 0$. Then for every $c \in \mathbb{C}$ the function $f_c(z) = f(z) - c$ satisfies*

$$N_+(f_c) - N_-(f_c) = \begin{cases} n_q - 1, & \text{if } n_p \leq n_q, \\ n_p, & \text{if } n_p > n_q + 1. \end{cases}$$

Proof. For each $c \in \mathbb{C}$ we can write f_c as in (1.2). Obviously, r and r_c have the same poles, and the argument given in the Introduction shows that f_c has finitely many zeros. Thus, if $\widetilde{M} > 0$ is sufficiently large, we have $P(f_c; B_{\widetilde{M}}(0)) = n_q$ and $N_{+/-}(f_c; B_{\widetilde{M}}(0)) = N_{+/-}(f_c)$.

First we assume $n_p \leq n_q$. Then $\lim_{|z|\to\infty} |r(z)|$ is finite (possibly zero), and for a sufficiently large $M \geq \widetilde{M} > 0$,

$$|f_c(z) + \overline{z}| = |r(z) - c| < |z| \leq |f_c(z)| + |\overline{z}| \quad \text{for all } z \in \partial B_M(0).$$

Using Theorem 2.2 and 2.3,

$$\begin{aligned} -1 &= V(-\overline{z}; \partial B_M(0)) = V(f_c; \partial B_M(0)) \\ &= N_+(f_c; B_M(0)) - N_-(f_c; B_M(0)) - P(f_c; B_M(0)) \\ &= N_+(f_c) - N_-(f_c) - n_q, \end{aligned}$$

which gives $N_+(f_c) - N_-(f_c) = n_q - 1$.

Next, we assume $n_p > n_q + 1$. Then r can be written as $r(z) = \widetilde{p}(z) + \widetilde{r}(z)$, where $\widetilde{p}$ is a polynomial of exact degree $n_p - n_q \geq 2$, and $\widetilde{r}$ is a rational function with $\lim_{|z|\to\infty} |\widetilde{r}(z)| = 0$. For a sufficiently large $M \geq \widetilde{M} > 0$,

$$|f_c(z) - \widetilde{p}(z)| = |\widetilde{r}(z) - \overline{z} - c| < |\widetilde{p}(z)| \leq |f_c(z)| + |\widetilde{p}(z)| \quad \text{for all } z \in \partial B_M(0).$$

Using again Theorem 2.2 and 2.3,

$$\begin{aligned} n_p - n_q &= V(\widetilde{p}; \partial B_M(0)) = V(f_c; \partial B_M(0)) \\ &= N_+(f_c; B_M(0)) - N_-(f_c; B_M(0)) - P(f_c; B_M(0)) \\ &= N_+(f_c) - N_-(f_c) - n_q, \end{aligned}$$

and hence $n_p = N_+(f_c) - N_-(f_c)$. □

Much of the proof of the next result is based on the proof of [1, Theorem C.3]. We nevertheless include all steps for clarity and completeness of our presentation.

Lemma 3.2. *Let f be as in (1.1). Then for every $c \in \mathbb{C}$ the function $f_c(z) = f(z) - c$ satisfies*

$$N_0(f_c) + N_-(f_c) \leq n_p + n_q - 1.$$

Proof. Let $z_0 \in \mathbb{C}$ be a non-sense-preserving zero of f, i.e., $r(z_0) = \overline{z}_0$ with $|r'(z_0)| \leq 1$. Then also $\overline{r}(r(z_0)) = z_0$, or

$$R(z_0) = z_0, \quad \text{where} \quad R(z) := (\overline{r} \circ r)(z).$$

The derivative of the rational function R is given by

$$R'(z) = (\overline{r} \circ r)'(z) = \overline{r}'(r(z))r'(z).$$

Thus,

$$|R'(z_0)| = |\overline{r}'(r(z_0))r'(z_0)| = |\overline{r}'(\overline{z}_0)r'(z_0)| = |\overline{r'(z_0)}r'(z_0)| = |r'(z_0)|^2 \leq 1,$$

which shows that z_0 is an attracting or rationally neutral fixed point of R. By Theorem 2.4, there exits a critical point z_c of R, i.e.,

$$R'(z_c) = \overline{r}'(r(z_c))r'(z_c) = 0, \tag{3.1}$$

such that

$$\lim_{k\to\infty} R^k(z_c) = z_0. \tag{3.2}$$

From (3.1) we obtain $r'(z_c) = 0$ or $r'(\overline{w}_c) = 0$, where $w_c := r(z_c)$. In the second case we use (3.2) and the continuity of $\overline{r}$ to obtain

$$\begin{aligned} \lim_{k\to\infty} R^k(\overline{w}_c) &= \lim_{k\to\infty} (\overline{r} \circ r)^k(\overline{w}_c) = \lim_{k\to\infty} (\overline{r} \circ r)^k(\overline{r}(\overline{z}_c)) \\ &= \overline{r}\Big(\lim_{k\to\infty} (r \circ \overline{r})^k(\overline{z}_c)\Big) = \overline{r}\Big(\overline{\lim_{k\to\infty} (\overline{r} \circ r)^k(z_c)}\Big) \\ &= \overline{r}\Big(\overline{\lim_{k\to\infty} R^k(z_c)}\Big) = \overline{r}(\overline{z}_0) = \overline{r(z_0)} = z_0. \end{aligned}$$

In summary, we have shown that if z_0 is a non-sense-preserving zero of f, and hence an attracting or rationally neutral fixed point of R, then there

exists a critical point $\widetilde{z}_c$ of r (in the first case $\widetilde{z}_c = z_c$, and in the second case $\widetilde{z}_c = \overline{w}_c$) with $\lim_{k\to\infty} R^k(\widetilde{z}_c) = z_0$. Clearly, different fixed points of R attract disjoint sets of (critical) points, and therefore $N_0(f) + N_-(f)$ is less than or equal to the number of zeros of

$$r'(z) = \frac{p'(z)q(z) - p(z)q'(z)}{q(z)^2},$$

which is at most $n_p + n_q - 1$. Finally, for every $c \in \mathbb{C}$ we can write f_c in the form (1.2). Now, by the same argument as above, $N_0(f_c) + N_-(f_c)$ is less than or equal to the number of zeros of $r'_c = r'$, which is at most $n_p + n_q - 1$. □

In order to control the behavior of singular functions we will also need the fact that a small constant perturbation does not reduce the number of sense-preserving zeros of f.

Lemma 3.3. *Let f be as in (1.1) and let $z_1, \dots, z_k$ be the sense-preserving zeros of f. Then $N_+(f) \le N_+(f_c)$ for all $c \in \mathbb{C}$ with sufficiently small $|c|$.*

Proof. Since f has finitely many zeros we can always find an $\varepsilon > 0$, such that f is sense-preserving on $B_\varepsilon(z_j)$ for $j = 1, \dots, k$, and $B_\varepsilon(z_j) \cap B_\varepsilon(z_\ell) = \varnothing$ for $j \neq \ell$, and

$$\delta := \min\left\{|f(z)| : z \in \cup_{j=1}^k \partial B_\varepsilon(z_j)\right\} > 0.$$

In particular, the condition $\delta > 0$ just means that none of the boundaries $\partial B_\varepsilon(z_j)$ contains a zero of f. By construction, for all $z \in \cup_{j=1}^k \partial B_\varepsilon(z_j)$ we have for $|c| < \delta$

$$|f(z) - f_c(z)| = |c| < \delta \le |f(z)| \le |f(z)| + |f_c(z)|.$$

With Theorem 2.3 we get, for each $j = 1, \dots, k$,

$$1 = N_+(f; B_\varepsilon(z_j)) = V(f; \partial B_\varepsilon(z_j)) = V(f_c; \partial B_\varepsilon(z_j)) = N_+(f_c; B_\varepsilon(z_j)),$$

where in the third equality we used that a constant shift preserves the orientation of f on $\mathbb{C}$. □

Proof of Theorem 1.1: Let f be as in (1.1) and let $c \in \mathbb{C}$ be arbitrary. Due to Lemma 2.1, there exists a sequence $\{c_k\}_{k\in\mathbb{N}} \subset \mathbb{C}$, such that the functions $f_{c_k}(z) := f(z) - c_k$ are regular, and $c_k \to c$. If f_c is regular, we can chose $c_k = c$ for all $k \in \mathbb{N}$.

If $n_p < n_q$, then for sufficiently small $|c_k - c|$,

$$\begin{aligned} N(f_c) &= N_+(f_c) + N_0(f_c) + N_-(f_c) \\ &\le N_+(f_{c_k}) - N_-(f_{c_k}) + N_-(f_{c_k}) + N_0(f_c) + N_-(f_c) \\ &\le (n_q - 1) + (n_p + n_q - 1) + (n_p + n_q - 1) \\ &= 2n_p + 3n_q - 3, \end{aligned}$$

where we have used Lemma 3.1–3.3. Analogously, if $n_p > n_q + 1$, then

$$\begin{aligned} N(f_c) &\le N_+(f_{c_k}) - N_-(f_{c_k}) + N_-(f_{c_k}) + N_0(f_c) + N_-(f_c) \\ &\le n_p + (n_p + n_q - 1) + (n_p + n_q - 1) \\ &= 3n_p + 2n_q - 2. \end{aligned}$$

Finally, if $n_p = n_q$, then the rational function r in (1.1) can be written as

$$r(z) = \frac{\alpha z^n + \widetilde{p}(z)}{z^n + \widetilde{q}(z)} = \frac{\widetilde{p}(z) - \alpha\widetilde{q}(z)}{z^n + \widetilde{q}(z)} + \alpha,$$

where $\alpha \neq 0$, $\deg(\widetilde{p}) < n_p = n$ and $\deg(\widetilde{q}) < n_q = n$. Now the numerator degree is

$$\widetilde{n}_p := \deg(\widetilde{p} - \alpha\widetilde{q}) \le n_p - 1,$$

and applying the bound from the first case to the function

$$f_{-\alpha}(z) := \frac{\widetilde{p}(z) - \alpha\widetilde{q}(z)}{z^n + \widetilde{q}(z)} - \overline{z} + \alpha$$

gives

$$N(f) = N(f_{-\alpha}) \le 2\widetilde{n}_p + 3n_q - 3 \le 5n_p - 5,$$

which completes the proof. □

We will now show that any function f as in (1.1) that attains one of the bounds of Theorem 1.1 is regular, which generalizes [10, Theorem 3.1].

Lemma 3.4. *Let f be as in* (1.1) *and suppose that f is singular. Then there exists a constant $c \in \mathbb{C}$ such that $N_+(f) < N_+(f_c)$.*

Proof. Note that f has at least one sense-preserving zero due to Lemma 3.1. Let $z_1, \dots, z_k$ be the sense-preserving zeros of f with the corresponding disks $B_\varepsilon(z_1), \dots, B_\varepsilon(z_k)$ as well as $\delta > 0$ as in the proof of Lemma 3.3.

Let z_0 be a singular zero of f. We then have $|f(z)| < \delta$ in $B_{\widetilde{\varepsilon}}(z_0)$ and

$$B_{\widetilde{\varepsilon}}(z_0) \cap B_\varepsilon(z_j) = \varnothing, \quad j = 1, \dots, k,$$

if $\widetilde{\varepsilon} > 0$ is small enough. Let $\widetilde{z} \in B_{\widetilde{\varepsilon}}(z_0)$ be arbitrary with $|r'(\widetilde{z})| > 1$. Then $|f(\widetilde{z})| < \delta$, and the proof of Lemma 3.3 shows that the function $\widetilde{f} := f - f(\widetilde{z})$ has one sense-preserving zero in each of the k disks $B_\varepsilon(z_1), \dots, B_\varepsilon(z_k)$. Moreover, $\widetilde{f}$ has an additional sense-preserving zero at $\widetilde{z}$, which means that $N_+(f) < N_+(\widetilde{f})$. □

Note that the bounds of Lemma 3.3 and 3.4 also hold for N_-.

Theorem 3.5. *Let f as in* (1.1) *and $c \in \mathbb{C}$. If f_c attains one of the bounds of Theorem 1.1, then f_c is regular.*

Proof. Let f be as in (1.1) with $n_p \le n_q$, and let $c \in \mathbb{C}$ be arbitrary. Due to Lemma 2.1 we can choose a sequence $\{c_k\}_{k\in\mathbb{N}}$, such that the functions f_{c_k} are regular and $c_k \to c$. Using Lemma 3.1–3.3 we obtain

$$\begin{aligned} N_+(f_c) &\le N_+(f_{c_k}) = N_+(f_{c_k}) - N_-(f_{c_k}) + N_-(f_{c_k}) \\ &\le (n_q - 1) + (n_p + n_q - 1) = n_p + 2n_q - 2. \end{aligned} \tag{3.3}$$

Now suppose that f_c attains the bound of Theorem 1.1. Then by Lemma 3.2 we get

$$\begin{aligned} 2n_p + 3n_q - 3 = N(f_c) &= N_+(f_c) + N_-(f_c) + N_0(f_c) \\ &\le N_+(f_{c_k}) + (n_p + n_q - 1), \end{aligned}$$

and with (3.3) we obtain $N_+(f_c) = n_p + 2n_p - 2$. If f_c would be singular, we could choose a constant $\widetilde{c} \in \mathbb{C}$ such that $N_+(f_{\widetilde{c}}) > N_+(f_c)$, but this is in contradiction to the upper bound (3.3), which holds for an arbitrary constant.

The proof for the case $n_p > n_q + 1$ is analogous. □

4 Discussion of Theorem 1.1

The reason why the special case $n_p = n_q + 1$ is "missing" in Theorem 1.1 is that this case is not covered in Lemma 3.1. Note that in this case we can write $r(z) = \alpha z + \widetilde{r}(z)$ for some $\alpha \neq 0$ and with $\lim_{|z|\to\infty} |\widetilde{r}(z)|$ finite (possibly zero). If $|\alpha| > 1$, then for a sufficiently large $M \ge \widetilde{M} > 0$ we obtain (cf. the proof of Lemma 3.1)

$$|f_c(z) - \alpha z| = |\widetilde{r}(z) - \overline{z} - c| < |\alpha z| \le |f_c(z)| + |\alpha z|$$

for all $z \in \partial B_M(0)$, leading to $1 = n_p - n_q = N_+(f_c) - N_-(f_c) - n_q$, or $N_+(f_c) - N_-(f_c) = n_q + 1 = n_p$. This gives the bound $N(f_c) \le 3n_p + 2n_q - 2 = 5n_p - 4$, but we know that $N(f_c) \le 5n_p - 6$ from [9]. For the case $|\alpha| = 1$ the method of proof used for Lemma 3.1 would give no result, and for $|\alpha| < 1$ we would indeed obtain $N(f_c) \le 5n_p - 6$, since then

$$|f_c(z) + \overline{z}| = |\widetilde{r}(z) + \alpha z - c| < |z| \le |f_c(z)| + |\overline{z}|$$

for all $z \in \partial B_M(0)$, giving $N_+(f_c) - N_-(f_c) = n_q - 1 = n_p - 2$.

Apart from the special case $n_p = n_q + 1$, Theorem 1.1 covers all possible choices of n_p and n_p with $n = \max\{n_p, n_q\} \ge 2$, and all previous results in this area that we are aware of. In particular:

(i) For any choices of n_p and n_q in Theorem 1.1 (except $n_p = n_q + 1$)[1] we get $N(f_c) \le 5n - 5$, which is the general bound from [4].

[1] The journal version of this article incorrectly stated "except $n_p = n_p + 1$".

(ii) For $n = n_p \geq 2$ and $n_q = 0$, Theorem 1.1 gives the bound for harmonic polynomials from [6].

(iii) For $n = n_p > n_q + 1$, Theorem 1.1 gives the same bound as in [7].

(iv) For $n = n_p = n_q + j$ with $j \geq 2$ we get $N(f) \leq 5n - 2(j+1)$. For $j = 2$ this is the same bound as in [10], and for $j > 2$ our new bound is smaller than the bound in [10] $(5n - 6)$.

(v) For $n = n_q = n_p + j$ with $j \geq 1$ we get $N(f) \leq 5n - (2j+3)$. For $j = 1$ this is the same bound as in [4], and for $j > 1$ our new bound is smaller than the previous one.

The following upper bounds on the maximal number of zeros of f as in (1.1) have been shown to be sharp:

$$N(f) \leq \begin{cases} 3n-2, & \text{if } (n_p, n_q) = (n, 0) \text{ [3]}, \\ 5n-5, & \text{if } (n_p, n_q) = (n-1, n) \text{ [11]}, \\ 5n-6, & \text{if } (n_p, n_q) = (n, n-1) \text{ [10]}. \end{cases}$$

Let f be the *Rhie function* from [11] (see also [12, 13]), which has a rational function r of the type $(n-1, n)$, and which has $5n-5$ zeros. The results in [8] imply that for sufficiently small $|c| > 0$ the function f_c has the same number of zeros as f. The corresponding rational function r_c then is of the type (n, n), so the bound $N(f) \leq 5n - 5$ is sharp also in the case (n, n). More generally, the sharpness of the bounds in Theorem 1.1 for $n_p \geq n_q$ is discussed in [7, Theorem C], while the case $n_p < n_q + 1$ remains a subject of future research.

Acknowledgements We thank an anonymous referee for several helpful suggestions, and in particular for pointing out the technical report [7].

References

[1] J. H. An and N. W. Evans, *The Chang–Refsdal lens revisited*, Monthly Notices of the Royal Astronomical Society, 369 (2006), pp. 317–334.

[2] L. Carleson and T. W. Gamelin, *Complex dynamics*, Universitext: Tracts in Mathematics, Springer-Verlag, New York, 1993.

[3] L. Geyer, *Sharp bounds for the valence of certain harmonic polynomials*, Proc. Amer. Math. Soc., 136 (2008), pp. 549–555.

[4] D. Khavinson and G. Neumann, *On the number of zeros of certain rational harmonic functions*, Proc. Amer. Math. Soc., 134 (2006), pp. 1077–1085.

[5] D. Khavinson and G. Neumann, *From the fundamental theorem of algebra to astrophysics: a "harmonious" path*, Notices Amer. Math. Soc., 55 (2008), pp. 666–675.

[6] D. Khavinson and G. Świątek, *On the number of zeros of certain harmonic polynomials*, Proc. Amer. Math. Soc., 131 (2003), pp. 409–414.

[7] S.-Y. Lee and N. Makarov, *Sharpness of connectivity bounds for quadrature domains*, arXiv:math/1411.3415v1, (2014).

[8] J. Liesen and J. Zur, *How constant shifts affect the zeros of certain rational harmonic functions*, arXiv:math/1702.07593v1, (2017). To appear in Comput. Methods Funct. Theory.

[9] R. Luce, O. Sète, and J. Liesen, *Sharp parameter bounds for certain maximal point lenses*, Gen. Relativity Gravitation, 46, Art. No. 1736 (2014), pp. 1–16.

[10] R. Luce, O. Sète, and J. Liesen, *A note on the maximum number of zeros of* $r(z) - \overline{z}$, Comput. Methods Funct. Theory, 15 (2015), pp. 439–448.

[11] S. H. Rhie, *n-point gravitational lenses with 5(n-1) images*, arXiv:astro-ph/0305166v1, (2003).

[12] O. Sète, R. Luce, and J. Liesen, *Creating images by adding masses to gravitational point lenses*, Gen. Relativity Gravitation, 47, Art. No. 42 (2015), pp. 1–8.

[13] O. Sète, R. Luce, and J. Liesen, *Perturbing rational harmonic functions by poles*, Comput. Methods Funct. Theory, 15 (2015), pp. 9–35.

A Newton method for harmonic mappings in the plane*

Olivier Sète† Jan Zur†

Abstract

We present an iterative root finding method for harmonic mappings in the complex plane, which is a generalization of Newton's method for analytic functions. The complex formulation of the method allows an analysis in a complex variables spirit. For zeros close to poles of $f = h + \overline{g}$ we construct initial points for which the harmonic Newton iteration is guaranteed to converge. Moreover, we study the number of solutions of $f(z) = \eta$ close to the critical set of f for certain $\eta \in \mathbb{C}$. We provide a MATLAB implementation of the method, and illustrate our results with several examples and numerical experiments, including phase plots and plots of the basins of attraction.

Keywords: Zeros of harmonic mappings, Newton's method, basins of attraction, domain coloring, Wilmshurst's conjecture, gravitational lensing.

Mathematics subject classification (2010): 30C55; 30D05; 31A05; 37F99; 65E05.

1 Introduction

We study harmonic mappings, i.e., functions with local decomposition

$$f(z) = h(z) + \overline{g(z)} \tag{1.1}$$

in the complex plane, where h and g are analytic. The function f itself is not analytic in general, as it consists of an analytic and an anti-analytic term.

The modern treatment of harmonic mappings started from the landmark paper [7] of Clunie and Sheil-Small about univalent harmonic mappings in the plane. See also the comprehensive textbook of Duren [11]. While [11]

*A final version of this preprint manuscript appeared as follows: O. Sète and J. Zur. A Newton method for harmonic mappings in the plane. *IMA J. Numer. Anal.* (2020), 40(4), pp. 2777–2801, doi:10.1093/imanum/drz042.

†TU Berlin, Institute of Mathematics, MA 3-3, Straße des 17. Juni 136, 10623 Berlin, Germany. {sete,zur}@math.tu-berlin.de

considers univalent harmonic mappings, also the multivalent case has been intensively studied; see e.g. the collection of open problems [5] by Bshouty and Lyzzaik. Problem 3.7 in [5] deals with the maximum number of zeros of harmonic polynomials $f(z) = p(z) + \overline{q(z)}$. Wilmshurst proved an upper bound for this number and conjectured another bound depending on $\deg(p)$ and $\deg(q)$ [45]; see Section 6.1 and the more recent publications concerning Wilmshurst's conjecture [24, 15, 19]. Khavinson and Świątek [23], and Geyer [13] settled the conjecture, including sharpness, for the special case $f(z) = p(z) - \overline{z}$. Khavinson and Neumann [21] generalized these results to rational harmonic functions $f(z) = r(z) - \overline{z}$; see also [28, 26]. The sharpness of the bound in the rational case was previously known due to the astrophysicist Rhie [36]. As it turns out, harmonic mappings model gravitational lensing, an astrophysical phenomenon, where, due to deflection by massive objects, a light source seems brightened, distorted or multiplied for an observer. We refer to the expository articles [22, 34], and the survey [3]. More recent articles on harmonic mappings with applications to gravitational lensing are [4, 20, 29, 37, 38, 27, 25].

The results about the zeros of harmonic mappings published so far are more theoretical, and we are not aware of any specialized method to compute them, which may explain the lack of (numerical) examples in the respective publications. In this paper we focus on the numerical computation of zeros of harmonic mappings with a Newton method.

The global behavior of Newton's method on $\mathbb{R}^2$, see e.g. [33, Ch. 7], has been less treated than the one of Newton's method for analytic or anti-analytic functions in $\mathbb{C}$. However, the recent publications [8, 9] give new impulses to this subject. Since harmonic mappings are generalizations of both analytic and anti-analytic functions, our work may also lead to new insights on the dynamics of Newton's method on $\mathbb{R}^2$. Our approach is not covered by the standard theory of complex dynamics [6, 30], nor by anti-holomorphic dynamics, see [32, 31], since the corresponding Newton map (3.7) is neither analytic nor harmonic.

The paper is organized as follows. In Section 2 we recall the definition and properties of harmonic mappings as well as Newton's method in Banach spaces. We then derive the harmonic Newton iteration and more generally a Newton iteration in complex notation for non-analytic but real differentiable functions in Section 3. We investigate zeros of $f = h + \overline{g}$ close to poles in Section 4. In particular, we construct initial points from the coefficients of the Laurent series of h and g, for which the harmonic Newton iteration is guaranteed to converge to zeros of f. In Section 5, we study solutions of $f(z) = \eta$, where z is close to a singular zero z_0 of f, and for certain small $\eta \in \mathbb{C}$. In Section 6, we consider several (numerical) examples. Finally, we give a summary and discuss possible future research in Section 7.

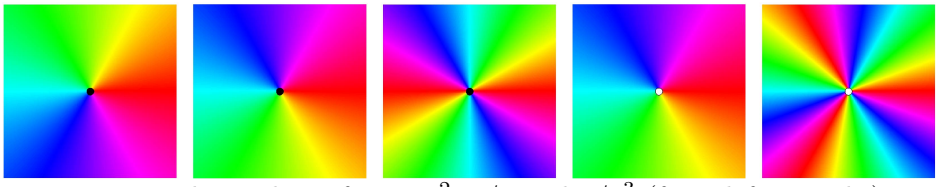


Figure 1: Phase plots of z, $\overline{z}$, z^2, $1/z$ and $1/\overline{z}^3$ (from left to right).

2 Mathematical background

In this section we recall properties of harmonic mappings and classical convergence results for Newton's method.

2.1 Harmonic mappings

The *Wirtinger derivatives* of a complex function f are

$$\partial_z f = \frac{1}{2}(\partial_x f - i\partial_y f), \quad \partial_{\overline{z}} f = \frac{1}{2}(\partial_x f + i\partial_y f),$$

where $z = x + iy$ with $x, y \in \mathbb{R}$; see [11, Section 1.2], [35, Section 1.4], or [44, p. 144]. They satisfy

$$\partial_x f = \partial_z f + \partial_{\overline{z}} f, \quad \partial_y f = i(\partial_z f - \partial_{\overline{z}} f), \quad \partial_{\overline{z}} \overline{f(z)} = \overline{\partial_z f(z)}. \tag{2.1}$$

A *harmonic mapping* is a function $f : \Omega \to \mathbb{C}$ defined on an open set $\Omega \subseteq \mathbb{C}$ and with

$$\Delta f = \partial_{xx} f + \partial_{yy} f = 4\partial_{\overline{z}}\partial_z f = 0.$$

Such a function has a local decomposition $f(z) = h(z) + \overline{g(z)}$, where h and g are analytic functions of z which are unique up to an additive constant [12, p. 412] or [11, p. 7]. The *Jacobian* of a harmonic mapping f at $z \in \Omega$ is

$$J_f(z) = |\partial_z f(z)|^2 - |\partial_{\overline{z}} f(z)|^2 = |h'(z)|^2 - |g'(z)|^2. \tag{2.2}$$

We call f *sense-preserving* (or orientation-preserving) at $z \in \Omega$ if $J_f(z) > 0$, *sense-reversing* if $J_f(z) < 0$, and *singular* if $J_f(z) = 0$, and similarly for zeros of f, e.g., z_0 is a singular zero of f if $J_f(z_0) = 0$. The points where f is singular form the *critical set*

$$\mathcal{C} = \{z \in \mathbb{C} : J_f(z) = 0\}. \tag{2.3}$$

One way to visualize complex functions are *phase plots*, where the domain is colored according to the phase $f(z)/|f(z)|$ of f. We use the standard color scheme described in [44] for all phase plots, and accordingly use white for the value ∞ (e.g., for poles of h or g) and black for zeros; see Figure 1. A comprehensive discussion of phase plots can be found in [44].

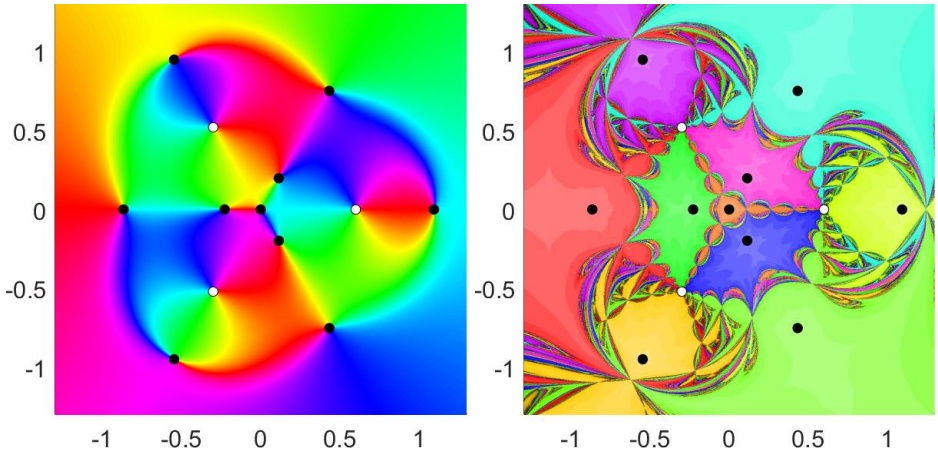


Figure 2: Left: Phase plot of $f(z) = z^2/(z^3 - 0.6^3) - \overline{z}$; see Example 2.1. White dots are the poles, black dots are the zeros. Right: Basins of attraction of the zeros of f in the harmonic Newton method; see Section 3 below.

Example 2.1. Consider the rational harmonic function

$$f(z) = h(z) + \overline{g(z)} = \frac{z^{n-1}}{z^n - r^n} - \overline{z}, \tag{2.4}$$

where h has the poles $re^{i2\pi k/n}$, $k = 0, 1, \ldots, n-1$. The function f has $3n+1$ zeros provided that $r > 0$ is sufficiently small; see [28] for a detailed analysis. Figure 2 shows the phase plot of f for $n = 3$ and $r = 0.6$, the 3 poles of f and the 10 zeros. Depending on the orientation, the phase plot of f near a zero looks like the phase plot of z (sense-preserving) or $\overline{z}$ (sense-reversing) near 0 in Figure 1.

2.2 Newton's method in Banach spaces

Let $F : D \to Y$, $D \subseteq X$ open, be a continuously Fréchet differentiable map between two Banach spaces X, Y. The Newton iteration with initial point $x_0 \in D$ is

$$x_{k+1} = x_k - F'(x_k)^{-1} F(x_k), \quad k \geq 0. \tag{2.5}$$

Under some regularity conditions, the sequence of Newton iterates $(x_k)_k$ is quadratically convergent ($\|x_{k+1} - x_*\| \leq \text{const} \cdot \|x_k - x_*\|^2$), if x_0 is sufficiently close to a zero of F; see e.g. [46, Prop. 5.1]. The next two theorems quantify "sufficiently close".

The Newton-Kantorovich theorem guarantees existence of a zero close to the initial point of the Newton iteration. In the following, $B(x_0; r)$ and $\overline{B}(x_0; r)$ denote the open and closed balls with center x_0 and radius r.

Theorem 2.2 (Newton-Kantorovich, [10, Thm. 2.1]). *Let $F : D \to Y$ be a continuously Fréchet differentiable map with $D \subseteq X$ open and convex. For an initial point $x_0 \in D$ let $F'(x_0)$ be invertible. Suppose that*

$$\|F'(x_0)^{-1}F(x_0)\| \leq \alpha,$$
$$\|F'(x_0)^{-1}(F'(y) - F'(x))\| \leq \omega_0\|y - x\| \quad \text{for all } x, y \in D.$$

Let $h_0 = \alpha\omega_0$ and $\rho = (1 - \sqrt{1 - 2h_0})/\omega_0$, and suppose that

$$h_0 \leq \frac{1}{2} \quad \text{and} \quad \overline{B}(x_0; \rho) \subseteq D.$$

Then the sequence $(x_k)_k$ of Newton iterates is well defined, remains in $\overline{B}(x_0; \rho)$, and converges to some x_* with $F(x_*) = 0$. If $h_0 < \frac{1}{2}$, then the convergence is quadratic.

The refined Newton-Mysovskii theorem guarantees uniqueness of a zero, but not its existence. Given a zero, it quantifies a neighborhood in which the Newton iteration will converge, as well as a constant for the quadratic convergence estimate.

Theorem 2.3 (refined Newton-Mysovskii theorem, [10, Thm. 2.3]). *Let $F : D \to \mathbb{R}^n$ be a continuously differentiable map with $D \subseteq \mathbb{R}^n$ open and convex. Suppose that $F'(x)$ is invertible for each $x \in D$. Suppose that*

$$\|F'(x)^{-1}(F'(y) - F'(x))(y - x)\| \leq \omega\|y - x\|^2 \quad \text{for all } x, y \in D. \tag{2.6}$$

Let $F(x) = 0$ have a solution $x_* \in D$. For the initial point x_0 suppose that $\overline{B}(x_*; \|x_0 - x_*\|) \subseteq D$ and that

$$\|x_0 - x_*\| < \frac{2}{\omega}.$$

Then the sequence of Newton iterates $(x_k)_k$ is well defined, remains in the open ball $B(x_*; \|x_0 - x_*\|)$, converges to x_*, and fulfills

$$\|x_{k+1} - x_*\| \leq \frac{\omega}{2}\|x_k - x_*\|^2.$$

Moreover, the solution x_* is unique in $B(x_*; 2/\omega)$.

We frequently use Landau's $\mathcal{O}$-notation for complex functions. By $f(z) + \mathcal{O}(z^k)$ we mean an expression $f(z)+\psi(z)$, where $\psi(z) \in \mathcal{O}(z^k)$, i.e., $|\psi(z)/z^k|$ is bounded from above by a constant (usually for $z \to 0$ or $z \to \infty$, depending on the context).

Remark 2.4. Using the Taylor expansion $\sqrt{1+z} = 1 + \frac{z}{2} + \mathcal{O}(z^2)$, we find the asymptotic expression

$$\rho = \frac{1 - \sqrt{1 - 2h_0}}{\omega_0} = \frac{1 - (1 - h_0 + \mathcal{O}(h_0^2))}{\omega_0} = \alpha + \mathcal{O}(\alpha h_0), \tag{2.7}$$

in Theorem 2.2, whenever the *Kantorovich quantity* h_0 is sufficiently small.

While we only use the above results, other convergence results for Newton's method in Banach spaces exist; see e.g. [39, 43].

3 The harmonic Newton method

We derive the harmonic Newton iteration and discuss its implementation.

3.1 The Newton iteration in the plane in complex notation

Let $f : \Omega \to \mathbb{C}$ be (continuously) differentiable with respect to $x = \mathrm{Re}(z)$ and $y = \mathrm{Im}(z)$, where we do not require for the moment that f is analytic or harmonic. We identify $\mathbb{C}$ with $\mathbb{R}^2$, and f with

$$F : \Omega \to \mathbb{R}^2, \quad F(z) = \begin{bmatrix} \mathrm{Re}(f(z)) \\ \mathrm{Im}(f(z)) \end{bmatrix}, \quad z = x + iy, \tag{3.1}$$

which is a smooth function of x and y.

Since Re and Im are $\mathbb{R}$-linear and continuous, they commute with ∂_x and ∂_y. Thus, $F'(z)$ has the matrix representation

$$F'(z) = \begin{bmatrix} \mathrm{Re}(\partial_z f(z)) + \mathrm{Re}(\partial_{\overline{z}} f(z)) & -\mathrm{Im}(\partial_z f(z)) + \mathrm{Im}(\partial_{\overline{z}} f(z)) \\ \mathrm{Im}(\partial_z f(z)) + \mathrm{Im}(\partial_{\overline{z}} f(z)) & \mathrm{Re}(\partial_z f(z)) - \mathrm{Re}(\partial_{\overline{z}} f(z)) \end{bmatrix}, \tag{3.2}$$

where we used the Wirtinger derivatives (2.1). The inverse of $F'(z)$ is

$$F'(z)^{-1} = \frac{1}{J_f(z)} \begin{bmatrix} \mathrm{Re}(\partial_z f(z)) - \mathrm{Re}(\partial_{\overline{z}} f(z)) & \mathrm{Im}(\partial_z f(z)) - \mathrm{Im}(\partial_{\overline{z}} f(z)) \\ -\mathrm{Im}(\partial_z f(z)) - \mathrm{Im}(\partial_{\overline{z}} f(z)) & \mathrm{Re}(\partial_z f(z)) + \mathrm{Re}(\partial_{\overline{z}} f(z)) \end{bmatrix},$$

provided $J_f(z) = |\partial_z f(z)|^2 - |\partial_{\overline{z}} f(z)|^2 \neq 0$; compare (2.2). We identify $F'(z)$ and its inverse with the $\mathbb{R}$-linear maps on $\mathbb{C}$

$$\begin{aligned} F'(z)(w) &= (\partial_z f(z))w + (\partial_{\overline{z}} f(z))\overline{w}, \\ F'(z)^{-1}(w) &= \frac{1}{J_f(z)}(\overline{\partial_z f(z)}w - (\partial_{\overline{z}} f(z))\overline{w}). \end{aligned}$$

Then the Newton iteration (2.5) for F can be rewritten in $\mathbb{C}$ as

$$z_{k+1} = z_k - \frac{\overline{\partial_z f(z_k)} f(z_k) - \partial_{\overline{z}} f(z_k)\overline{f(z_k)}}{|\partial_z f(z_k)|^2 - |\partial_{\overline{z}} f(z_k)|^2}, \quad k \geq 0. \tag{3.3}$$

3.2 The harmonic Newton iteration

For a harmonic mapping with local decomposition $f = h + \overline{g}$, we have $\partial_z f(z) = h'(z)$ and $\partial_{\overline{z}} f(z) = \partial_{\overline{z}} \overline{g(z)} = \overline{g'(z)}$ with (2.1), so that $F'(z)$ and $F'(z)^{-1}$ can be identified with

$$F'(z)(w) = h'(z)w + \overline{g'(z)w}, \tag{3.4}$$

$$F'(z)^{-1}(w) = \frac{1}{J_f(z)}(\overline{h'(z)}w - \overline{g'(z)w}), \tag{3.5}$$

and the Newton iteration (3.3) takes the form

$$z_{k+1} = z_k - \frac{\overline{h'(z_k)}f(z_k) - \overline{g'(z_k)f(z_k)}}{|h'(z_k)|^2 - |g'(z_k)|^2}, \quad k \geq 0, \tag{3.6}$$

which we call the *harmonic Newton iteration*. It reduces to the classical Newton iteration $z_{k+1} = z_k - f(z_k)/f'(z_k)$ when f is analytic ($f = h$). Similarly, if f is anti-analytic ($f = \overline{g}$), the iteration simplifies to $z_{k+1} = z_k - g(z_k)/g'(z_k)$. Hence, the harmonic Newton iteration (3.6) is a natural generalization of the classical Newton iteration.

We define the *harmonic Newton map*

$$H_f : \Omega \setminus \mathcal{C} \to \mathbb{C}, \quad z \mapsto H_f(z) = z - \frac{\overline{h'(z)}f(z) - \overline{g'(z)f(z)}}{|h'(z)|^2 - |g'(z)|^2}, \tag{3.7}$$

which is in general neither analytic nor harmonic. Note that $z_* \in \Omega \setminus \mathcal{C}$ is a fixed point of H_f if and only if z_* is a non-singular zero of f. Borrowing notation from complex dynamics, see e.g. [6, 30], we define the *basin of attraction* of a zero z_* of f as

$$A(z_*) = \{z \in \mathbb{C} : \lim_{k\to\infty} H_f^k(z) = z_*\},$$

where $H_f^1 = H_f$ and $H_f^k = H_f \circ H_f^{k-1}$ for $k \geq 2$. Note that the basin of attraction of a non-singular zero z_* contains an open neighborhood of z_*, by the local convergence of Newton's method. For singular zeros this needs not be the case; see Example 6.3.

To visualize the dynamics of the harmonic Newton map, we use the following standard domain coloring technique; see, e.g., [14, 42]. We color every point according to which basin it belongs to. The color level indicates the number of iterations, the darker the more iterations were required. Points where the harmonic Newton iteration does not converge are colored in black; see Figure 2.

3.3 Implementation

Following Higham [17, Sect. 25.5], we stop the harmonic Newton iteration when the residual $|f(z_k)|$ or the relative difference between two iterates $|z_{k+1} - z_k|/|z_{k+1}|$ are less than given tolerances, or when a prescribed maximum number of steps has been performed. This gives the *harmonic Newton method*; see Algorithm 1. Note that, when $f(z_*) = 0$ and $|z_k - z_*|$ is sufficiently small, we have $|z_{k+1} - z_*| \leq 2c|z_{k+1} - z_k|^2$, where c is the constant from the (local) quadratic convergence of the Newton method; see [17, Sect. 25.5].

Our MATLAB implementation of Algorithm 1 is displayed in Figure 3. To apply a harmonic Newton step not only to a single point, but to a vector or matrix of points, we vectorize the iteration (3.6), see lines 25–26. This

Algorithm 1 HARMONIC NEWTON METHOD

Input: f, h', g', and $z_0 \in \mathbb{C}$, $N \in \mathbb{N}$, $\varepsilon_f, \varepsilon_z > 0$

Output: approximated zero of f

1: **for** $k = 0, \ldots, N-1$ **do**

2: $\quad z_{k+1} = z_k - \dfrac{\overline{h'(z_k)} f(z_k) - \overline{g'(z_k) f(z_k)}}{|h'(z_k)|^2 - |g'(z_k)|^2}$

3: $\quad$ **if** $|f(z_{k+1})| < \varepsilon_f$ **or** $|z_{k+1} - z_k| < \varepsilon_z |z_{k+1}|$ **then**

4: $\quad\quad$ **return** z_{k+1}

5: $\quad$ **end if**

6: **end for**

7: **return** z_N

also requires vectorized function handles for f, h', g'. Vectorization is particularly useful when we have several initial points, since applying the method to a matrix of initial points is usually faster than applying it to each point individually. When all points are iterated simultaneously, some might already have converged while others have not. However, further iterating a point that has numerically converged can be harmful in finite precision, e.g., when a point is close to the critical set, a further Newton step may lead to division by zero. Hence, we stop the iteration individually for each point, by removing them from the list of active points; see lines 15 and 33. Note that computing the next iterate in (3.6) requires only the three function evaluations $f(z_k)$, $h'(z_k)$ and $g'(z_k)$.

Computing the next iterate by (3.6) or by solving the 2×2 real linear algebraic system corresponding to

$$F'(z_k)(z_{k+1} - z_k) = -F(z_k) \tag{3.8}$$

with (3.2) are equivalent in exact arithmetic. This is not the case in finite precision, where the Jacobian $J_f(z)$ can be numerically zero, although z is not on the critical set. To avoid division by zero, it is preferable to compute the next iterate by solving (3.8), instead of inverting $F'(z_k)$ explicitly. However, we use (3.6) in our numerical experiments (except in Example 4.1), see line 2 in Algorithm 1 and lines 25–26 in Figure 3, because (3.6) can be easily vectorized.

4 Finding zeros close to poles

The complex formulation (3.6) makes the harmonic Newton iteration amenable to analysis in a complex variables spirit. We investigate zeros of harmonic mappings $f = h + \overline{g}$ close to poles z_0, i.e., points where $\lim_{z \to z_0} |f(z)| = \infty$; see [40, Def. 2.1]. More precisely, we construct initial points from the Laurent coefficients of h and g, for which the harmonic Newton iteration is guaranteed to converge to zeros of f. We start with an example.

```
function [z, numiter] = harmonicNewton(f, dh, dg, z, maxit, restol, steptol)

% Set default values:
if ((nargin < 5) || isempty(maxit)), maxit = 50; end
if ((nargin < 6) || isempty(restol)), restol = 1e-14; end
if ((nargin < 7) || isempty(steptol)), steptol = 1e-14; end
% Setup
active = 1:numel(z);                    % track which points to iterate
numiter = (maxit+1)*ones(size(z));      % number of iterations
% Check initial points:
fz = f(z);
converged = (abs(fz) < restol);         % already converged
diverged = (isnan(fz) | isinf(fz));     % NaN or Inf
numiter(converged) = 0;                 % no Newton steps required
active(converged | diverged) = [];      % do not further iterate
fz = fz(active);
% Harmonic Newton iteration:
for kk = 1:maxit
    if (isempty(active)), return, end   % every point has converged
    zold = z(active);
    % Evaluate functions:
    dhz = dh(zold);
    dgz = dg(zold);
    % Newton step:
    z(active) = zold ...
        - (conj(dhz).*fz - conj(dgz.*fz))./(abs(dhz).^2 - abs(dgz).^2);
    % Convergence check:
    fz = f(z(active));
    converged = (abs(fz) < restol) | ...
        (abs(z(active) - zold) < steptol*abs(zold));
    diverged = (isnan(fz) | isinf(fz)); % NaN or inf
    numiter(active(converged)) = kk;    % took kk iterations
    active(converged | diverged) = [];  % remove converged points
    fz(converged | diverged) = [];      % remove converged points
end
end
```

Figure 3: MATLAB implementation of the harmonic Newton method used in the examples.

Example 4.1. We consider $f(z) = \tan(z) - \overline{z}$; see Figure 4. The function has a singular zero at the origin, and one (real) zero close to each pole of the tangent, except $\pm\frac{\pi}{2}$. To compute the zeros of f we apply the harmonic Newton method to a grid of initial points in $[-8, 8] \times [-2, 2]$ with mesh size 0.2, tolerances 10^{-14}, and with up to 50 steps. The residual at our computed zeros satisfies $|f(z_k)| \leq 1.5600 \cdot 10^{-13}$.

Since the Jacobian $J_f(z) = |1 + \tan(z)^2|^2 - 1$ is numerically zero for $|z| \leq 10^{-9}$, we use (3.8) instead of (3.6) to compute the basins of attraction in this example. We observe that the basin of 0 is huge, while the basins of zeros close to poles are small. Moreover, the size of the basins decreases for

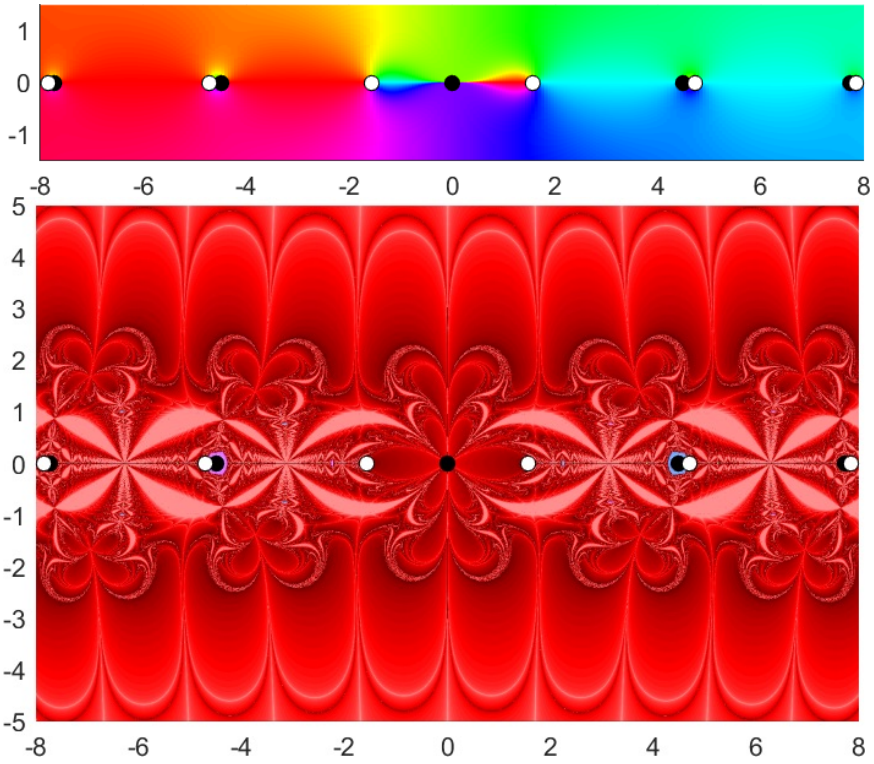


Figure 4: The function $f(z) = \tan(z) - \overline{z}$ from Example 4.1. Top: Phase plot, zeros (black dots), poles (white dots). Bottom: Basins of attraction.

larger poles, and the poles seem to be on the boundary of the basins.

Note that the black points on the imaginary axis are due to a numerically vanishing Jacobian, as above. Analytically, the harmonic Newton map is

$$H_f(iy) = i\left(y - \frac{y + \tanh(y)}{1 + \operatorname{sech}(y)^2}\right) \quad \text{for } y \in \mathbb{R},$$

and one can prove that every point on the imaginary axis converges to 0.

Next, we construct initial points for which the harmonic Newton iteration converges to zeros close to a pole. For this, we need the following lemma.

Lemma 4.2. *The equation $az + \overline{bz} = c$ has a unique solution if and only if $|a| \neq |b|$. In that case, the unique solution is $z = \frac{\overline{a}c - \overline{b}\overline{c}}{|a|^2 - |b|^2}$.*

Proof. Splitting $az + \overline{bz} = c$ in its real and imaginary part yields a real 2×2 linear algebraic system with determinant $|a|^2 - |b|^2$, from which we obtain the assertion. □

To apply the Newton-Kantorovich theorem to the harmonic Newton iteration (3.6), we identify $\mathbb{C}$ with $\mathbb{R}^2$ as in Section 3.1. First, we derive bounds

for the constants α and ω_0 in Theorem 2.2 using the formulas for F' and $(F')^{-1}$ from (3.4) and (3.5). We find

$$|F'(z_0)^{-1}(w)| \le \frac{|h'(z_0)| + |g'(z_0)|}{\big||h'(z_0)|^2 - |g'(z_0)|^2\big|}|w| = \frac{|w|}{\big||h'(z_0)| - |g'(z_0)|\big|},$$

and so

$$|F'(z_0)^{-1}(F(z_0))| \le \frac{|f(z_0)|}{\big||h'(z_0)| - |g'(z_0)|\big|} \le \alpha \tag{4.1}$$

and

$$\begin{aligned}\|F'(z_0)^{-1}(F'(y) - F'(x))\| &= \max_{|z|=1}|F'(z_0)^{-1}(F'(y) - F'(x))(z)| \\ &\le \frac{\max_{|z|=1}(|h'(y) - h'(x)| + |g'(y) - g'(x)|)|z|}{\big||h'(z_0)| - |g'(z_0)|\big|} \\ &\le \frac{\sup_{t\in D}|h''(t)| + \sup_{t\in D}|g''(t)|}{\big||h'(z_0)| - |g'(z_0)|\big|}|y - x| \\ &\le \omega_0|y - x|. \end{aligned} \tag{4.2}$$

Next we present the main result of this section.

Theorem 4.3. *Let $f = h + \overline{g}$, where*

$$h(z) = \sum_{k=-n}^{\infty} a_k(z - z_0)^k, \quad g(z) = \sum_{k=-n}^{\infty} b_k(z - z_0)^k,$$

with $n \ge 1$, and $|a_{-n}| \ne |b_{-n}|$. Suppose that $c = -(a_0 + \overline{b}_0) \ne 0$, and let $z_1, \ldots, z_n$ be the n solutions of

$$(z - z_0)^n = \frac{|a_{-n}|^2 - |b_{-n}|^2}{\overline{a}_{-n}c - \overline{b}_{-n}\overline{c}}. \tag{4.3}$$

We then have for sufficiently large $|c|$:

1. *There exist n distinct zeros of f near z_0.*
2. *The zeros in 1. are the limits of the harmonic Newton iteration with initial points $z_1, \ldots, z_n$.*

Proof. We apply the Newton-Kantorovich theorem for each of the n initial points z_j. Without loss of generality, we assume $z_0 = 0$. Then "$|c|$ sufficiently large" is equivalent to "$|z_j|$ sufficiently small".

The cases $n = 1$ and $n \ge 2$ differ slightly, and we begin with $n \ge 2$. First, we determine α from (4.1). With (4.3) and Lemma 4.2 we obtain[1]

$$a_{-n}(z_j - z_0)^{-n} + \overline{b_{-n}(z_j - z_0)^{-n}} = c = -(a_0 + \overline{b}_0),$$

[1]In the journal version of this article, the second term of the following equation is incorrectly $\overline{b}_{-n}(z_j - z_0)^{-n}$.

and

$$|f(z_j)| \le (|a_{-n+1}| + |b_{-n+1}|)|z_j|^{-n+1} + \mathcal{O}(|z_j|^{-n+2}), \tag{4.4}$$

and from

$$h'(z) = -na_{-n}z^{-n-1} + \mathcal{O}(z^{-n}), \quad g'(z) = -nb_{-n}z^{-n-1} + \mathcal{O}(z^{-n}),$$

also

$$\big||h'(z_j)| - |g'(z_j)|\big| = n\big||a_{-n}| - |b_{-n}|\big||z_j|^{-n-1} + \mathcal{O}(|z_j|^{-n}),$$

so that

$$\begin{aligned}\frac{|f(z_j)|}{\big||h'(z_j)| - |g'(z_j)|\big|} &\le \frac{(|a_{-n+1}| + |b_{-n+1}|)|z_j|^{-n+1} + \mathcal{O}(|z_j|^{-n+2})}{n\big||a_{-n}| - |b_{-n}|\big||z_j|^{-n-1} + \mathcal{O}(|z_j|^{-n})} \\ &= \frac{|a_{-n+1}| + |b_{-n+1}|}{n\big||a_{-n}| - |b_{-n}|\big|}|z_j|^2 + \mathcal{O}(|z_j|^3) = \alpha.\end{aligned}$$

Here we used that $1/(1 + \mathcal{O}(x)) = 1 + \mathcal{O}(x)$ for $x \to 0$.

Next we determine ω_0 from (4.2). Let $D = \{z \in \mathbb{C} : |z - z_j| < q|z_j|\}$ for some $0 < q < 1$. From

$$h''(z) = n(n+1)a_{-n}z^{-n-2} + \mathcal{O}(z^{-n-1}),$$

we get

$$\sup_{t\in D}|h''(t)| \le n(n+1)|a_{-n}|((1-q)|z_j|)^{-n-2} + \mathcal{O}(|z_j|^{-n-1}),$$

and similarly for g. Then

$$\begin{aligned}&\frac{\sup_{t\in D}|h''(t)| + \sup_{t\in D}|g''(t)|}{\big||h'(z_j)| - |g'(z_j)|\big|} \\ &\le \frac{n(n+1)(|a_{-n}| + |b_{-n}|)((1-q)|z_j|)^{-n-2} + \mathcal{O}(|z_j|^{-n-1})}{n\big||a_{-n}| - |b_{-n}|\big||z_j|^{-n-1} + \mathcal{O}(|z_j|^{-n})} \\ &= \frac{(n+1)(|a_{-n}| + |b_{-n}|)(1-q)^{-n-2}}{\big||a_{-n}| - |b_{-n}|\big|}|z_j|^{-1} + \mathcal{O}(1) = \omega_0,\end{aligned}$$

and $h_0 = \alpha\omega_0 = \mathcal{O}(|z_j|)$. Hence, for any $0 < q < 1$, we have $h_0 < \frac{1}{2}$ for sufficiently small $|z_j|$.

It remains to show that $\rho < q|z_j|$. Using (2.7) we get

$$\rho = \alpha + \mathcal{O}(\alpha h_0) = \mathcal{O}(|z_j|^2) < q|z_j| \tag{4.5}$$

for sufficiently small $|z_j|$. By Theorem 2.2, the sequence of harmonic Newton iterates remains in the closed disk $\overline{D}(z_j;\rho)$ and converges to a zero of f.

Finally, we show that the n disks $\overline{D}(z_j;\rho)$, $j = 1, 2, \ldots, n$, are disjoint. By construction, the z_j are n-th roots. Thus, they have equispaced angles,

satisfy $|z_j - z_{j+1}| = 2\sin(\pi/n)|z_j|$, and have distance $|z_j|$ from the origin. Since $\rho = \mathcal{O}(|z_j|^2)$, the disks are disjoint for sufficiently small $|z_j|$. This concludes the proof in the case $n \geq 2$.

For $n = 1$, the proof differs in (4.4), where we have $|f(z_1)| = \mathcal{O}(|z_1|)$ (instead of $\mathcal{O}(1)$). Hence, $\alpha = \mathcal{O}(|z_1|^3)$ and the proof can be completed as for $n \geq 2$. We omit the details. □

Remark 4.4.

1. The assumption "$|c|$ is sufficiently large" is necessary to guarantee the existence of a zero close to the pole; see Figure 4, where we have no zero "close" to the poles $\pm\frac{\pi}{2}$ and where $c = \pm\frac{\pi}{2}$, respectively.

2. One of the coefficients a_{-n} and b_{-n} in Theorem 4.3 may be zero (but not both). In particular, it is possible that only one of the functions h and g has a pole at z_0.

3. The case $|a_{-n}| = |b_{-n}|$ is excluded in Theorem 4.3. Such functions can have nonisolated zeros, e.g., $f(z) = z^{-n} + \overline{z}^{-n} = 2\,\mathrm{Re}(z^{-n})$.

4. The Laurent coefficients of h and g are given by contour integrals, and can be computed numerically by quadrature, e.g. with the trapezoidal rule; see, e.g., [41]. To compute several Laurent coefficients at once, we can use the trapezoidal rule in combination with the DFT; see [1, Sect. 3.8.1], or [16, Sect. 13.4].

By the transformation $z \mapsto \frac{1}{z}$, we obtain the following corollary for zeros "close to infinity".

Corollary 4.5. *Let $f = h + \overline{g}$, where*

$$h(z) = \sum_{k=-\infty}^{n} a_k z^k, \quad g(z) = \sum_{k=-\infty}^{n} b_k z^k, \quad \text{for } |z| > R > 0,$$

with $n \geq 1$, and $|a_n| \neq |b_n|$. Furthermore, let $c = -(a_0 + \overline{b}_0) \neq 0$ and let $z_1, \ldots, z_n$ be the n solutions of

$$z^n = \frac{\overline{a}_n c - \overline{b}_n \overline{c}}{|a_n|^2 - |b_n|^2}.$$

We then have for sufficiently large $|c|$:

1. *There exist n distinct zeros of f "close to infinity".*

2. *The zeros in 1. are the limits of the harmonic Newton iteration with initial points $z_1, \ldots, z_n$.*

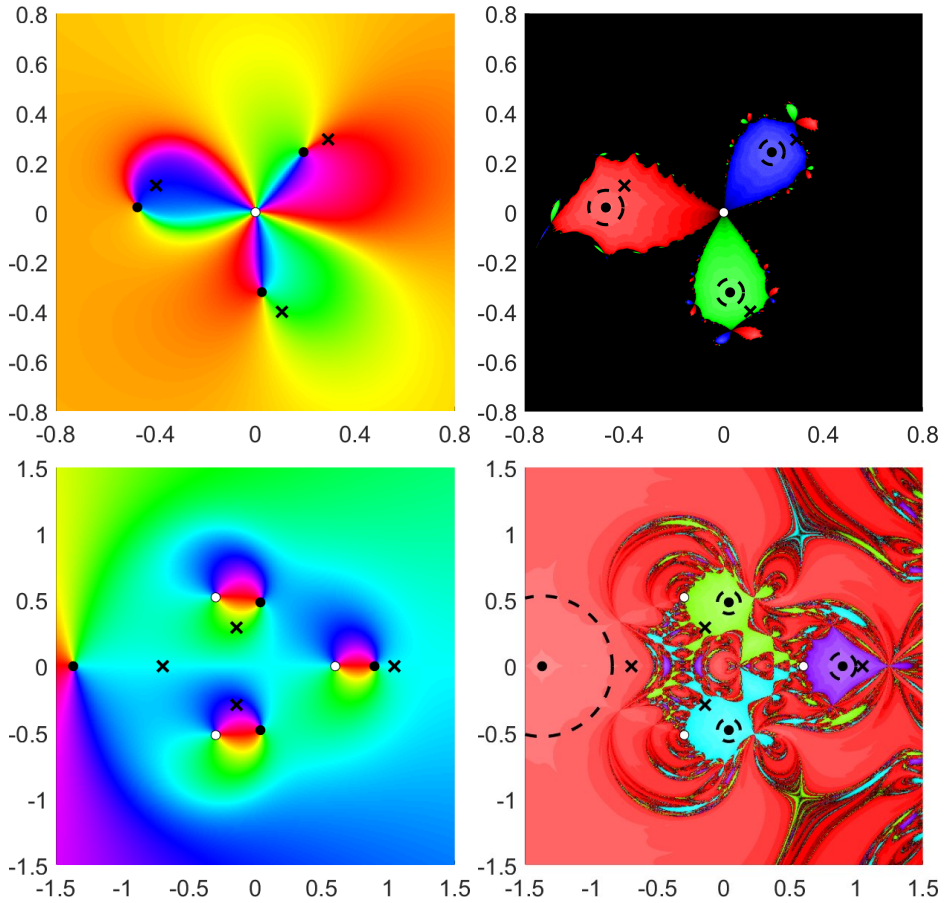


Figure 5: Phase plots (left) and basins of attraction (right) of $f(z) = \frac{1}{z^3} + 10(1+i) + \overline{\left(\frac{2i}{z^2}\right)}$ (top) and $f(z) = \frac{z^2}{z^3 - 0.6^3} - 0.7 - \overline{z}$ (bottom) from Example 4.6. Crosses indicate the initial points for the harmonic Newton method from Theorem 4.3 and Corollary 4.5. The dashed circles are regions of convergence guaranteed by Theorem 2.3.

Example 4.6. Let $f(z) = \frac{1}{z^3} + 10(1+i) + \overline{\left(\frac{2i}{z^2}\right)}$, where the analytic and anti-analytic part have poles of order 3 and 2 in $z_0 = 0$, with $c = -10(1+i)$. As predicted by Theorem 4.3, there are three distinct zeros close to $z_0 = 0$; see Figure 5. The initial points from the theorem (marked by crosses) lie close to the edges of the basins. After 9 steps of the harmonic Newton iteration, the maximal residual is $3.5527 \cdot 10^{-15}$. For points in the black region, the sequence of harmonic Newton iterates tends to ∞. Note that the boundary of the basins of attraction seems to have fractal character; compare to the "Douady rabbit".

Next, we consider the function $f(z) = \frac{z^2}{z^3 - 0.6^3} - 0.7 - \overline{z}$, which is the function from Example 2.1 with an additive perturbation. The function has three simple poles, one zero close to each pole, and a fourth zero; see

Figure 5. The crosses indicate the initial points from Theorem 4.3 and from Corollary 4.5 (for the fourth zero). As for the previous example, they are close to the boundary of the basins of attraction of the zeros, which again suggests that the constant c cannot be much smaller. After 7 steps of the harmonic Newton iteration, the maximal residual of the iterates is $4.3299 \cdot 10^{-15}$.

The plot of the basins of attraction of the two functions also show dashed circles, which are regions of convergence guaranteed by the refined Newton-Mysovskii theorem (Theorem 2.3). For a zero z_* of f, this region is computed as follows. Since the theorem guarantees convergence for initial points in a disk, we choose D as a disk with radius r and center z_*. To estimate the constant ω, we evaluate (2.6) on a grid in D. The points in the disk with radius $\min\{r, 2/\omega\}$ and center z_* are guaranteed to converge to z_*, and we maximize over r to compute the largest such disk. We observe that our initial points are not covered by Theorem 2.3, but the iterates converge nevertheless to zeros of the functions.

5 Behavior near the critical set

The harmonic Newton map is not defined on the critical set $\mathcal{C}$; see Section 3. Nevertheless, in Examples 4.1 and 6.3 below, the harmonic Newton method approximates singular zeros, i.e., $f(z_0) = 0$ with $z_0 \in \mathcal{C}$, reasonably well. We now consider solutions of $f(z) = \eta$ with small $|\eta|$ and z close to the singular zero z_0. For rational harmonic functions $f(z) = r(z) - \overline{z}$, these solutions play a crucial role in the study of the (global) number of zeros of f; see [25]. As it turns out, the number of (local) solutions of $f(z) = \eta$ close to z_0 can be $0, 1, 2$, or 3, depending on η and z_0.

Let f be a general harmonic mapping with a singular zero z_0. We prove existence of solutions of the equation $f(z) = \eta$ by showing convergence of the harmonic Newton iteration with suitable initial points to zeros of $f(z) - \eta$. Depending on η, the shape of the basins of attraction and the number of solutions can change dramatically.

For simplicity we transform $f = h + \overline{g}$ in a local normal form. Let $z_0 \in \mathcal{C}$ be a singular zero of f with $|h'(z_0)| = |g'(z_0)| \neq 0$; see (2.3). We then have

$$f(z) = h(z) + \overline{g(z)} = \sum_{k=0}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=0}^{\infty} b_k (z - z_0)^k}$$

with $0 = f(z_0) = a_0 + \overline{b}_0$, and $0 \neq \overline{b}_1 = a_1 e^{i2\theta}$, $\theta \in [0, \pi[$. Hence,

$$f(z) = \frac{\overline{b}_1}{e^{i\theta}} \left(\sum_{k=1}^{\infty} \frac{a_k}{\overline{b}_1} e^{i(k+1)\theta} (e^{-i\theta}(z - z_0))^k + \overline{\sum_{k=1}^{\infty} \frac{b_k}{b_1} e^{i(k-1)\theta} (e^{-i\theta}(z - z_0))^k} \right),$$

and substituting

$$\zeta = e^{-i\theta}(z - z_0), \quad \alpha_k = \frac{a_k}{\overline{b}_1} e^{i(k+1)\theta} = \frac{a_k}{a_1} e^{i(k-1)\theta}, \quad \beta_k = \frac{b_k}{b_1} e^{i(k-1)\theta}, \tag{5.1}$$

gives $f(z) = \overline{b}_1 e^{-i\theta} \widetilde{f}(\zeta)$, where $\widetilde{f}$ is the unique *local normal form* of f at z_0,

$$\widetilde{f}(\zeta) = \zeta + \overline{\zeta} + \sum_{k=2}^{\infty} \alpha_k \zeta^k + \overline{\sum_{k=2}^{\infty} \beta_k \zeta^k}. \tag{5.2}$$

Here we used translation ($\zeta_0 = 0$), and scaling and rotation ($\alpha_1 = \beta_1 = 0$).

We take solutions of the truncated problem $\zeta + \overline{\zeta} + \alpha_2 \zeta^2 + \overline{\beta}_2 \overline{\zeta}^2 = \eta$ that are close to the origin as initial points for the harmonic Newton method applied to $\widetilde{f}(\zeta) - \eta$.

Lemma 5.1. *Let $f = h + \overline{g}$ be as in* (5.2),

$$f(\zeta) = \zeta + \overline{\zeta} + \sum_{k=2}^{\infty} \alpha_k \zeta^k + \overline{\sum_{k=2}^{\infty} \beta_k \zeta^k}.$$

In particular, f has a singular zero at 0 with $|h'(0)| = |g'(0)| \neq 0$. Furthermore, let $f_{\delta c}(\zeta) = f(\zeta) - \delta c$ with $c = -(\alpha_2 + \overline{\beta}_2)$ and real $\delta \in \mathbb{R}$.

1. *For $\operatorname{Im}(c) \neq 0$ and sufficiently small $\delta > 0$, the function $f_{\delta c}$ has two distinct zeros close to 0. These zeros are the limits of the harmonic Newton iteration for $f_{\delta c}$ with the initial points $\zeta_\pm = \pm i\sqrt{\delta}$.*
2. *For $\operatorname{Im}(c) = 0$, $|\alpha_2| \neq |\beta_2|$, and sufficiently small $|\delta|$, the harmonic Newton iteration for $f_{\delta c}$ with initial point $\zeta_3 = \frac{1-\sqrt{1-\delta c^2}}{c}$ converges to a zero of $f_{\delta c}$.*

Proof. To apply the Newton-Kantorovich theorem to $f_{\delta c}$ and each initial point, we estimate α and ω_0; see (4.1) and (4.2).

Let $\operatorname{Im}(c) \neq 0$. Since $\zeta_\pm = \pm i\sqrt{\delta}$ solves $\zeta + \overline{\zeta} + \alpha_2 \zeta^2 + \overline{\beta}_2 \overline{\zeta}^2 = \delta c$, we have

$$|f_{\delta c}(\zeta_\pm)| \leq (|\alpha_3| + |\beta_3|)|\zeta_\pm|^3 + \mathcal{O}(|\zeta_\pm|^4).$$

From

$$h'(\zeta) = 1 + 2\alpha_2 \zeta + \mathcal{O}(\zeta^2), \quad g'(\zeta) = 1 + 2\beta_2 \zeta + \mathcal{O}(\zeta^2),$$

we have, using $\overline{\zeta}_\pm = -\zeta_\pm$,

$$|h'(\zeta_\pm)|^2 = h'(\zeta_\pm)\overline{h'(\zeta_\pm)} = 1 + 4i \operatorname{Im}(\alpha_2)\zeta_\pm + \mathcal{O}(|\zeta_\pm|^2),$$

and similarly for g. Thus

$$\begin{aligned} \big||h'(\zeta_\pm)| - |g'(\zeta_\pm)|\big| &= \frac{\big||h'(\zeta_\pm)|^2 - |g'(\zeta_\pm)|^2\big|}{|h'(\zeta_\pm)| + |g'(\zeta_\pm)|} = \frac{4|\operatorname{Im}(\alpha_2 - \beta_2)\zeta_\pm| + \mathcal{O}(|\zeta_\pm|^2)}{2 + \mathcal{O}(|\zeta_\pm|)} \\ &= 2|\operatorname{Im}(\alpha_2 - \beta_2)||\zeta_\pm| + \mathcal{O}(|\zeta_\pm|^2) \\ &= 2|\operatorname{Im}(c)||\zeta_\pm| + \mathcal{O}(|\zeta_\pm|^2), \end{aligned}$$

for sufficiently small $\delta > 0$. Together we find

$$\begin{aligned}\frac{|f_{\delta c}(\zeta_\pm)|}{\big||h'(\zeta_\pm)| - |g'(\zeta_\pm)|\big|} &\le \frac{(|\alpha_3| + |\beta_3|)|\zeta_\pm|^3 + \mathcal{O}(|\zeta_\pm|^4)}{2|\mathrm{Im}(c)||\zeta_\pm| + \mathcal{O}(|\zeta_\pm|^2)} \\ &= \frac{|\alpha_3| + |\beta_3|}{2|\mathrm{Im}(c)|}|\zeta_\pm|^2 + \mathcal{O}(|\zeta_\pm|^3) = \alpha.\end{aligned}$$

Next, we estimate ω_0 on $D_\pm = \{\zeta \in \mathbb{C} : |\zeta - \zeta_\pm| < q|\zeta_\pm|\}$ for a $0 < q < 1$. From

$$h''(\zeta) = 2\alpha_2 + \mathcal{O}(\zeta), \quad g''(\zeta) = 2\beta_2 + \mathcal{O}(\zeta),$$

we have

$$\sup_{\tau\in D_\pm} |h''(\tau)| + \sup_{\tau\in D_\pm} |g''(\tau)| \le 2(|\alpha_2| + |\beta_2|) + \mathcal{O}(|\zeta_\pm|),$$

and

$$\begin{aligned}\frac{\sup_{\tau\in D_\pm}|h''(\tau)| + \sup_{\tau\in D_\pm}|g''(\tau)|}{\big||h'(\zeta_\pm)| - |g'(\zeta_\pm)|\big|} &\le \frac{2(|\alpha_2| + |\beta_2|) + \mathcal{O}(|\zeta_\pm|)}{2|\mathrm{Im}(c)||\zeta_\pm| + \mathcal{O}(|\zeta_\pm|^2)} \\ &= \frac{|\alpha_2| + |\beta_2|}{|\mathrm{Im}(c)||\zeta_\pm|} + \mathcal{O}(1) = \omega_0.\end{aligned}$$

Thus

$$h_0 = \alpha\omega_0 = \frac{(|\alpha_2| + |\beta_2|)(|\alpha_3| + |\beta_3|)}{2|\mathrm{Im}(c)|^2}|\zeta_\pm| + \mathcal{O}(|\zeta_\pm|^2),$$

so that $h_0 < \frac{1}{2}$ for sufficiently small δ, since $\zeta_\pm = \pm i\sqrt{\delta} \to 0$ for $\delta \to 0$. Finally, we show $\rho < q|\zeta_\pm|$. Using (2.7) we find that

$$\rho = \alpha + \mathcal{O}(\alpha h_0) = \frac{|\alpha_3| + |\beta_3|}{2|\mathrm{Im}(c)|}|\zeta_\pm|^2 + \mathcal{O}(|\zeta_\pm|^3) < q|\zeta_\pm|, \tag{5.3}$$

provided $\zeta_\pm$ is sufficiently small. Then, by Theorem 2.2, the sequence of harmonic Newton iterates remains in the closed disks $\overline{D}(\zeta_\pm; \rho)$ and converges to a zero of $f_{\delta c}$. Furthermore, both disks are disjoint, which implies that the iterations converge to two distinct zeros of $f_{\delta c}$.

For the second part, assume that $\mathrm{Im}(c) = 0$ and $|\alpha_2| \neq |\beta_2|$. The latter implies that $c \neq 0$. The only difference to the first case is in the estimate of $\big||h'(\zeta_3)| - |g'(\zeta_3)|\big|$. Since c is real, also $\zeta_3 = \frac{1-\sqrt{1-\delta c^2}}{c} = \frac{c}{2}\delta + \mathcal{O}(\delta^2)$ is real for $\delta \le \frac{1}{c^2}$, and we find from $h'(\zeta) = 1 + 2\alpha_2\zeta + \mathcal{O}(\zeta^2)$ that

$$|h'(\zeta_3)|^2 = h'(\zeta_3)\overline{h'(\zeta_3)} = 1 + 4\,\mathrm{Re}(\alpha_2)\zeta_3 + \mathcal{O}(|\zeta_3|^2),$$

and similarly for g. Then

$$\begin{aligned}\big||h'(\zeta_3)| - |g'(\zeta_3)|\big| = \frac{\big||h'(\zeta_3)|^2 - |g'(\zeta_3)|^2\big|}{|h'(\zeta_3)| + |g'(\zeta_3)|} &= \frac{|4\,\mathrm{Re}(\alpha_2 - \beta_2)\zeta_3 + \mathcal{O}(\zeta_3^2)|}{2 + \mathcal{O}(|\zeta_3|)} \\ &= 2|\mathrm{Re}(\alpha_2 - \beta_2)||\zeta_3| + \mathcal{O}(|\zeta_3|^2).\end{aligned}$$

The assumptions $|\alpha_2| \neq |\beta_2|$ and $\mathrm{Im}(c) = 0$ guarantee that $\mathrm{Re}(\alpha_2 - \beta_2) \neq 0$ and thus that $\big||h'(\zeta_3)| - |g'(\zeta_3)|\big| = \mathcal{O}(|\zeta_3|)$. We then obtain α, ω_0 in the same orders of magnitude as in the first case, so that $h_0 = \alpha\omega_0 = \mathcal{O}(|\zeta_3|) < \frac{1}{2}$ and $\rho < q|\zeta_3|$ for sufficiently small $|\delta|$. Therefore, the harmonic Newton iteration with initial point ζ_3 converges to a zero of $f_{\delta c}$. □

Theorem 5.2. *Let the harmonic mapping*

$$f(z) = h(z) + \overline{g(z)} = \sum_{k=1}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=1}^{\infty} b_k (z - z_0)^k}$$

have a singular zero at z_0 with $|a_1| = |b_1| \neq 0$. Furthermore, let $\theta \in [0, \pi[$ be defined by $\overline{b}_1 = a_1 e^{i2\theta}$, and let

$$f_{\delta c}(z) = f(z) - \delta c, \quad \text{with} \quad c = \overline{b}_1 e^{-i\theta} \widetilde{c} \quad \text{and} \quad \widetilde{c} = -\frac{a_2}{a_1} e^{i\theta} - \overline{\left(\frac{b_2}{b_1} e^{i\theta}\right)},$$

where δ is real.

1. *For $\mathrm{Im}(\widetilde{c}) \neq 0$ and sufficiently small $\delta > 0$, the function $f_{\delta c}$ has two distinct zeros close to z_0. These zeros are the limits of the harmonic Newton iteration for $f_{\delta c}$ with the initial points*

$$z_{\pm} = z_0 \pm i\sqrt{\delta} e^{i\theta} = z_0 \pm i\sqrt{\delta \frac{\overline{b}_1}{a_1}}.$$

2. *For $\mathrm{Im}(\widetilde{c}) = 0$, $|a_2| \neq |b_2|$, and sufficiently small $|\delta|$, the harmonic Newton iteration with initial point*

$$z_3 = z_0 + \frac{1 - \sqrt{1 - \delta\widetilde{c}^2}}{\widetilde{c}} e^{i\theta},$$

converges to a zero of $f_{\delta c}$.

Proof. We transform f to its local normal form $\widetilde{f}$ at z_0, recall (5.1), (5.2), and $f(z) = \overline{b}_1 e^{-i\theta} \widetilde{f}(\zeta)$. Then $f(z) - \delta c = \overline{b}_1 e^{-i\theta} (\widetilde{f}(\zeta) - \delta\widetilde{c})$ with

$$\widetilde{c} = -\frac{a_2}{a_1} e^{i\theta} - \overline{\left(\frac{b_2}{b_1} e^{i\theta}\right)} = -(\alpha_2 + \overline{\beta}_2).$$

Let $\mathrm{Im}(\widetilde{c}) \neq 0$. Then, by Lemma 5.1, the harmonic Newton iteration with initial points $\zeta_{\pm} = \pm i\sqrt{\delta}$ converges to zeros of $\widetilde{f}_{\delta\widetilde{c}}$, when $\delta > 0$ is sufficiently small. Back transformation gives the desired result; see (5.1).

Let $\mathrm{Im}(\widetilde{c}) = 0$, and $|\alpha_2| \neq |\beta_2|$, which is equivalent to $|a_2| \neq |b_2|$. Then, by Lemma 5.1, the harmonic Newton iteration with initial point $\zeta_3 = \frac{1 - \sqrt{1 - \delta\widetilde{c}^2}}{\widetilde{c}}$ converges to a zero of $\widetilde{f}_{\delta\widetilde{c}}$, when $\delta > 0$ is sufficiently small. Back transformation gives the desired result. □

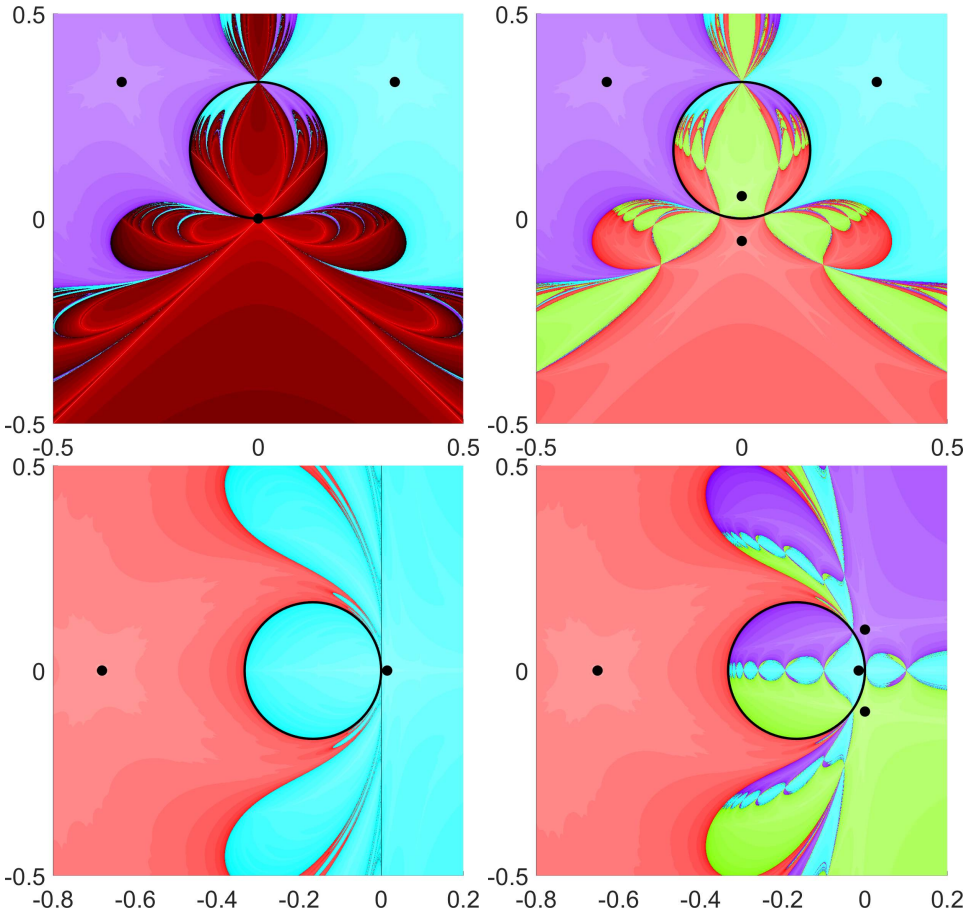


Figure 6: Basins of attraction. Top: $f_{-i\delta}(z) = z + \overline{z} + 2iz^2 + \overline{iz^2} + i\delta$ with $\delta = 0$ (left) and $\delta = 0.003$ (right). Bottom: $f_{-3\delta}(z) = z + \overline{z} + 2z^2 + \overline{z}^2 + 3\delta$ with $\delta = -0.01$ (left) and $\delta = 0.01$ (right). Black lines mark the critical sets.

Varying δ can lead to a bifurcation of the solutions of $f(z) = \delta c$, and the basins of attraction can merge or split; see Figure 6.

Remark 5.3.

1. The conditions $\mathrm{Im}(c) \neq 0$ in Lemma 5.1 and $\mathrm{Im}(\widetilde{c}) \neq 0$ in Theorem 5.2 imply that $f(z_0)$ is not a *cusp* of the *caustics* of f; see [25].

2. In Figure 6, the singular zero at $z = 0$ (top left) splits into two non-singular zeros (top right). Note that the harmonic Newton iteration converges quicker to the non-singular zeros than to the singular zero.

3. For f in Figure 6 (bottom left) we have $H_f(iy) = \frac{i}{2}(y - \frac{0.01}{y})$, $y \in \mathbb{R}$. Hence, the harmonic Newton iteration diverges for initial points on the imaginary axis, which explains the black line.

4. Figure 6 (bottom right) suggests that we have zeros of $f_{\delta c}$ close to $z_\pm = \pm i\sqrt{\delta}$ also in the case $\mathrm{Im}(c) = 0$ and $\delta > 0$. However, without further improvements, the above proof strategy breaks down for $z_\pm$ as initial points.

6 Further Examples

We illustrate the harmonic Newton method with further examples, using the MATLAB implementation in Figure 3.

6.1 Harmonic polynomials

Harmonic polynomials are of the form $f = p + \overline{q}$, where p and q are (analytic) polynomials with respective degrees n and m. Wilmshurst conjectured in [45] that the number of zeros of f for $n > m$ fulfills

$$N(f) \le 3n - 2 + m(m-1), \tag{6.1}$$

and proved it for $m = n - 1$, including sharpness of the bound. For $n = m$, Wilmshurst showed the bound $N(f) \le n^2$, provided that $\lim_{z\to\infty} f(z) = \infty$ holds. Otherwise f could have infinitely many zeros; compare Lemma 4.2. Later, the case $m = 1$ was settled in [23], and sharpness of the bound in [13]. For several other values of $1 < m < n - 1$, the conjecture was shown to be wrong; see [24, 15, 19].

Example 6.1. We consider the original example of Wilmshurst in [45] showing sharpness of (6.1) for $m = n - 1$. Let

$$f(z) = p(z) + \overline{q(z)} = z^n + (z-1)^n + \overline{i(z-1)^n - iz^n}, \tag{6.2}$$

which has $\deg(p) = n$, and $\deg(q) = n - 1$, and n^2 zeros.

To compute the zeros of f with the harmonic Newton method, we use a grid of initial points with mesh size 0.05 in the plot region in Figure 7. On the left we have the phase plots, and on the right the basins of attraction for $n = 3$ and the $n^2 = 9$ zeros (top), as well as for $n = 10$ with its 100 zeros (bottom). The plots are centered at $z = 0.5$, since $z \mapsto p(z + 0.5)$ and $z \mapsto q(z + 0.5)$ are even or odd, depending on n.

For $n = 3$, the maximal residual of f at the computed zeros is $9.8625 \cdot 10^{-15}$. For $n = 10$, we have $|f(z_j)| \le 3.0146 \cdot 10^{-14}$ at computed zeros with $|z_j - 0.5| \le 1$, and $|f(z_j)| \le 1.3738 \cdot 10^{-7}$ at zeros with $|z_j - 0.5| > 1$. The latter rather large absolute residual is an effect of floating point arithmetic. Indeed, the magnitude of z^n and $(z-1)^n$ is of order 10^8 at the outer zeros, which results in a loss of accuracy of about 8 digits when evaluating f. Evaluating the polynomials p and q with Horner's scheme [17, Sect. 5.1] does not alleviate this problem, since we still need to compute $p + \overline{q}$. However, the maximum of the relative residuals $|f(z_j)|/|z_j - 0.5|^{10}$ is $1.4010 \cdot 10^{-14}$, which is quite satisfactory.

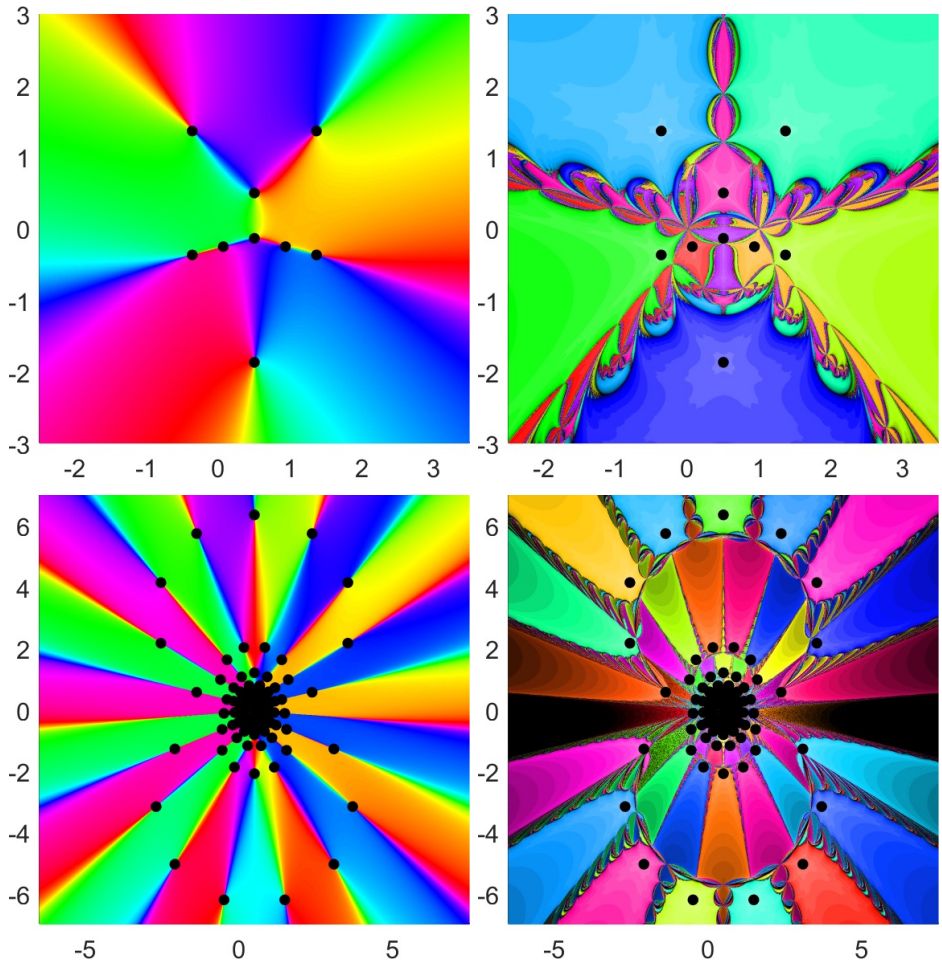


Figure 7: Phase plots (left) and basins of attraction (right) for Wilmshurst's harmonic polynomial (6.2) for $n = 3$ (top) and $n = 10$ (bottom).

6.2 Gravitational lensing

Gravitational lensing can be modeled with harmonic mappings, where positions of lensed images are zeros of these functions; see [22, 34, 3]. We consider point mass lenses and isothermal gravitational lenses.

Example 2.1 already featured a rational harmonic function from gravitational point mass models. Rhie [36] constructed from this example a gravitational lens with the maximum number of lensed images, or equivalently a rational harmonic function with the maximum number of zeros, showing sharpness of the bound in [21].

Example 6.2. Rhie's function

$$f(z) = (1-\varepsilon)\frac{z^{n-1}}{z^n - r^n} + \frac{\varepsilon}{z} - \overline{z} \tag{6.3}$$

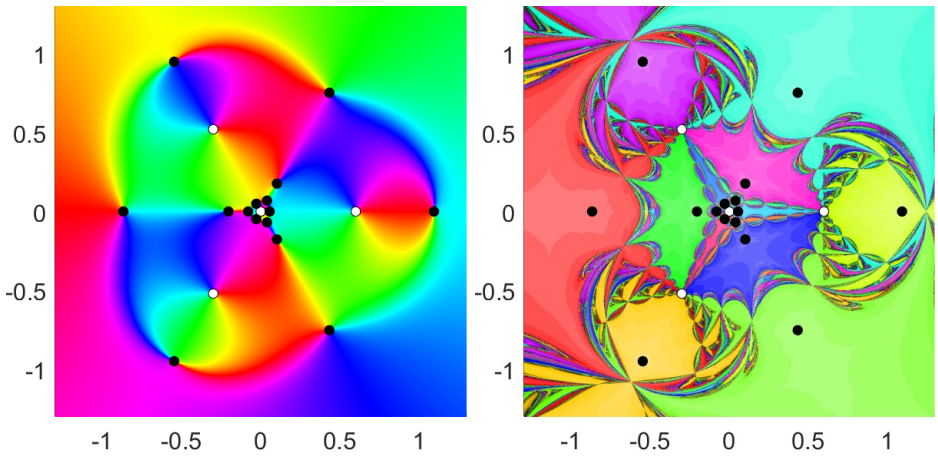


Figure 8: Phase plot (left) and basins of attraction (right) for Rhie's function (6.3) with $r = 0.6$, $n = 3$, and $\varepsilon = 0.004$; compare Figure 2.

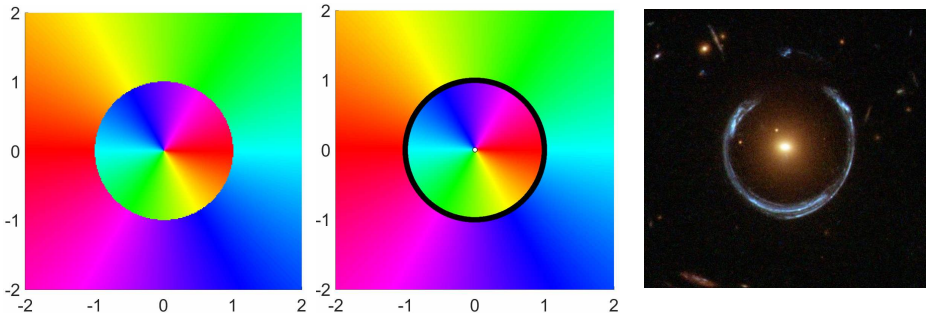


Figure 9: Phase plots of $f(z) = 1/z - \overline{z}$. The middle panel shows the computed zeros of f; see Example 6.3. The right panel shows an actual gravitional lens (image credit: ESA/Hubble & NASA).

is obtained from (2.4) by adding a pole at the origin (and normalizing the mass to be one). For sufficiently small $\varepsilon > 0$, it has $5n$ zeros; for a rigorous proof and quantification of the parameters ε and r see [28]. To compute the zeros of f with $n = 3$, $r = 0.6$ and $\varepsilon = 0.004$, we apply the harmonic Newton method to a grid of initial points (mesh size 0.05); see Figure 8. The maximal residual at the computed zeros is $9.9371 \cdot 10^{-15}$.

Example 6.3. A point mass lens with a single mass is known as the Chang-Refsdal lens and produces an Einstein ring; see [2]. It can be modeled by $f(z) = \frac{1}{z} - \overline{z}$. The zero set of f is the whole unit circle, in particular the zeros are not isolated. Moreover, each zero is singular. To compute zeros of f, we take a grid of initial points in $[-2, 2]^2$ (mesh size 0.02). Figure 9 shows the iterates after at most 30 steps (the mean number of steps is 6.4). The residuals satisfy $|f(z_k)| \leq 2.4887 \cdot 10^{-14}$. The harmonic Newton iteration for f simplifies to $z_{k+1} = \frac{2}{1+|z_k|^2} z_k$ for $|z_k| \neq 1$, showing that $z_k \to e^{i\varphi}$ if

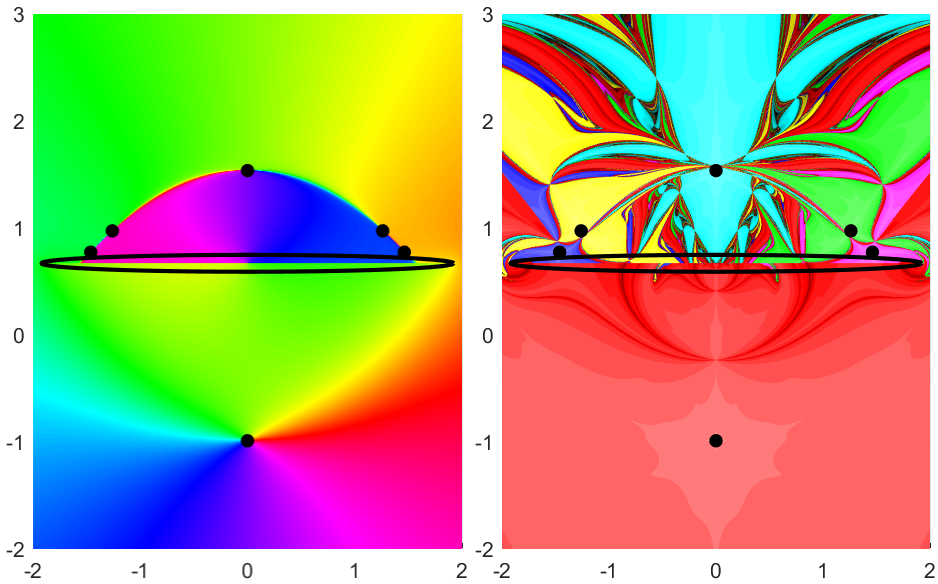


Figure 10: Phase plot (left) and basins of attraction (right) for the function $f(z) = z - \arcsin(1.92/\overline{z - 0.67i})$ in Example 6.4; compare to [20, Fig. 3].

$z_0 = r_0 e^{i\varphi} \neq 0$, i.e., the basin of attraction of $e^{i\varphi}$ is

$$A(e^{i\varphi}) = \{z = re^{i\varphi} : r > 0, r \neq 1\}.$$

Note that $A(e^{i\varphi})$ does not contain an open subset around $e^{i\varphi}$, which is due to the fact that the zeros of f are not isolated.

Example 6.4. We consider gravitational lensing by an isothermal elliptical galaxy with compactly supported mass; see [3, 20, 4]. After a change of variables, the lensed images are the zeros of the transcendental harmonic mapping

$$f(z) = z - \arcsin\left(\frac{k}{\overline{z} + \overline{w}}\right), \tag{6.4}$$

where w is the position of the light source (projected onto the lens plane), and k is a real constant related to the shape of the galaxy; see [20]. We take the principal branch of arcsine and compute the zeros of f by applying the harmonic Newton method to a grid of initial points in $[-2, 2]^2$ (mesh size 0.02). Figure 10 shows the phase plot of f (left) and the basins of attraction (right) for $k = 1.92$ and $w = -0.67i$ from [4]. The black ellipse is the shape of the galaxy, and the sharp edge in the phase plot (inside the ellipse) is the branch cut of arcsine.

7 Summary and Outlook

We derived a complex formulation of Newton's method for harmonic mappings and, more generally, for real differentiable functions in $\mathbb{C}$. The iterations (3.3) and its harmonic pendant (3.6) allow to compute the zeros of harmonic and even non-analytic complex functions without losing any advantages of Newton's method. The complex formulation makes the iteration amenable to analysis in a complex variables spirit. In particular, close to a pole of a harmonic mapping f, we derived initial points for which the harmonic Newton iteration is guaranteed to converge to zeros of f. Similarly, we derived initial points to obtain solutions of $f(z) = \eta$ for certain small $|\eta|$ and z close to a singular zero of f. In both cases, the convergence proof relies on the Newton-Kantorovich theorem. In particular, our theorems also show existence of zeros.

The harmonic Newton method should prove useful in the further study of harmonic polynomials, in particular related to Wilmshurst's conjecture, and more generally of harmonic mappings. Visualizing the basins of attraction can give new insights into the behavior of the functions.

For certain classes of harmonic mappings, e.g., harmonic polynomials, it could be interesting to find a set $\mathcal{S}$, such that for any zero z_* of f, the harmonic Newton iteration converges to z_* for at least one initial point in $\mathcal{S}$. This approach is related to Theorem 4.3 and Theorem 5.2. In the seminal paper [18], such a set $\mathcal{S}$ is constructed for complex (analytic) polynomials.

A further analysis of the harmonic Newton map H_f in the spirit of complex dynamics should be of great interest, e.g., points of indeterminacy, and the continuation of H_f (and of f itself) to the Riemann sphere. Both have not been addressed in this paper.

For analytic functions, many other iterative root finding methods exist; see e.g. [14, 42]. It could be interesting to generalize them to harmonic mappings. As a drawback we may have to introduce higher order derivatives for harmonic mappings. The efficiency index of several of these methods for analytic functions are compared in [42]. In this light, Newton's method should be a decent choice.

Acknowledgments. We thank Jörg Liesen for helpful comments on the manuscript. Moreover, we thank the anonymous referees for several helpful suggestions which lead to improvements of the presentation.

References

[1] M. J. Ablowitz and A. S. Fokas, *Complex variables: introduction and applications*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2nd ed., 2003.

[2] J. H. An and N. W. Evans, *The Chang–Refsdal lens revisited*, Monthly Notices Roy. Astronom. Soc., 369 (2006), pp. 317–334.

[3] C. Bénéteau and N. Hudson, *A survey on the maximal number of solutions of equations related to gravitational lensing*, in Complex analysis and dynamical systems, Trends Math., Birkhäuser/Springer, Cham, 2018, pp. 23–38.

[4] W. Bergweiler and A. Eremenko, *On the number of solutions of a transcendental equation arising in the theory of gravitational lensing*, Comput. Methods Funct. Theory, 10 (2010), pp. 303–324.

[5] D. Bshouty and A. Lyzzaik, *Problems and conjectures in planar harmonic mappings*, J. Anal., 18 (2010), pp. 69–81.

[6] L. Carleson and T. W. Gamelin, *Complex dynamics*, Universitext: Tracts in Mathematics, Springer-Verlag, New York, 1993.

[7] J. Clunie and T. Sheil-Small, *Harmonic univalent functions*, Ann. Acad. Sci. Fenn. Ser. A I Math., 9 (1984), pp. 3–25.

[8] R. De Leo, *"Simple Dynamics" conjectures for some real Newton maps on the plane*, arXiv:math/1812.00270v1, (2018).

[9] R. De Leo, *Julia sets of Newton maps of real quadratic polynomial maps on the plane*, arXiv:math/1812.11595v2, (2019).

[10] P. Deuflhard, *Newton methods for nonlinear problems*, vol. 35 of Springer Series in Computational Mathematics, Springer, Heidelberg, 2011.

[11] P. Duren, *Harmonic mappings in the plane*, vol. 156 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2004.

[12] P. Duren, W. Hengartner, and R. S. Laugesen, *The argument principle for harmonic functions*, Amer. Math. Monthly, 103 (1996), pp. 411–415.

[13] L. Geyer, *Sharp bounds for the valence of certain harmonic polynomials*, Proc. Amer. Math. Soc., 136 (2008), pp. 549–555.

[14] W. J. Gilbert, *Generalizations of Newton's method*, Fractals, 9 (2001), pp. 251–262.

[15] J. D. Hauenstein, A. Lerario, E. Lundberg, and D. Mehta, *Experiments on the zeros of harmonic polynomials using certified counting*, Exp. Math., 24 (2015), pp. 133–141.

[16] P. Henrici, *Applied and computational complex analysis. Vol. 3*, John Wiley & Sons, Inc., New York, 1986.

[17] N. J. Higham, *Accuracy and stability of numerical algorithms*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 2002.

[18] J. Hubbard, D. Schleicher, and S. Sutherland, *How to find all roots of complex polynomials by Newton's method*, Invent. Math., 146 (2001), pp. 1–33.

[19] D. Khavinson, S.-Y. Lee, and A. Saez, *Zeros of harmonic polynomials, critical lemniscates, and caustics*, Complex Anal. Synerg., 4, Art. No. 2 (2018), pp. 1–20.

[20] D. Khavinson and E. Lundberg, *Transcendental harmonic mappings and gravitational lensing by isothermal galaxies*, Complex Anal. Oper. Theory, 4 (2010), pp. 515–524.

[21] D. Khavinson and G. Neumann, *On the number of zeros of certain rational harmonic functions*, Proc. Amer. Math. Soc., 134 (2006), pp. 1077–1085.

[22] D. Khavinson and G. Neumann, *From the fundamental theorem of algebra to astrophysics: a "harmonious" path*, Notices Amer. Math. Soc., 55 (2008), pp. 666–675.

[23] D. Khavinson and G. Świątek, *On the number of zeros of certain harmonic polynomials*, Proc. Amer. Math. Soc., 131 (2003), pp. 409–414.

[24] S.-Y. Lee, A. Lerario, and E. Lundberg, *Remarks on Wilmshurst's theorem*, Indiana Univ. Math. J., 64 (2015), pp. 1153–1167.

[25] J. Liesen and J. Zur, *How constant shifts affect the zeros of certain rational harmonic functions*, Comput. Methods Funct. Theory, 18 (2018), pp. 583–607.

[26] J. Liesen and J. Zur, *The maximum number of zeros of $r(z) - \overline{z}$ revisited*, Comput. Methods Funct. Theory, 18 (2018), pp. 463–472.

[27] R. Luce and O. Sète, *The index of singular zeros of harmonic mappings of anti-analytic degree one*, Oberwolfach Report OWP 2017-03, (2017). To appear in Complex Var. Elliptic Equ.

[28] R. Luce, O. Sète, and J. Liesen, *Sharp parameter bounds for certain maximal point lenses*, Gen. Relativity Gravitation, 46, Art. No. 1736 (2014), pp. 1–16.

[29] R. Luce, O. Sète, and J. Liesen, *A note on the maximum number of zeros of $r(z) - \overline{z}$*, Comput. Methods Funct. Theory, 15 (2015), pp. 439–448.

[30] J. Milnor, *Dynamics in one complex variable*, vol. 160 of Annals of Mathematics Studies, Princeton University Press, Princeton, NJ, third ed., 2006.

[31] S. Mukherjee, S. Nakane, and D. Schleicher, *On multicorns and unicorns II: bifurcations in spaces of antiholomorphic polynomials*, Ergodic Theory Dynam. Systems, 37 (2017), pp. 859–899.

[32] S. Nakane and D. Schleicher, *On multicorns and unicorns. I. Antiholomorphic dynamics, hyperbolic components and real cubic polynomials*, Internat. J. Bifur. Chaos Appl. Sci. Engrg., 13 (2003), pp. 2825–2844.

[33] H.-O. Peitgen and P. H. Richter, *The beauty of fractals. Images of complex dynamical systems*, Springer-Verlag, Berlin, 1986.

[34] A. O. Petters, *Gravity's action on light*, Notices Amer. Math. Soc., 57 (2010), pp. 1392–1409.

[35] R. Remmert, *Theory of complex functions*, vol. 122 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1991.

[36] S. H. Rhie, *n-point gravitational lenses with 5(n-1) images*, arXiv:astro-ph/0305166v1, (2003).

[37] O. Sète, R. Luce, and J. Liesen, *Creating images by adding masses to gravitational point lenses*, Gen. Relativity Gravitation, 47, Art. No. 42 (2015), pp. 1–8.

[38] O. Sète, R. Luce, and J. Liesen, *Perturbing rational harmonic functions by poles*, Comput. Methods Funct. Theory, 15 (2015), pp. 9–35.

[39] S. Smale, *Newton's method estimates from data at one point*, in The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985), Springer, New York, 1986, pp. 185–196.

[40] T. J. Suffridge and J. W. Thompson, *Local behavior of harmonic mappings*, Complex Variables Theory Appl., 41 (2000), pp. 63–80.

[41] L. N. Trefethen and J. A. C. Weideman, *The exponentially convergent trapezoidal rule*, SIAM Rev., 56 (2014), pp. 385–458.

[42] J. L. Varona, *Graphic and numerical comparison between iterative methods.*, Math. Intell., 24 (2002), pp. 37–46.

[43] X. Wang, *Convergence of Newton's method and inverse function theorem in Banach space*, Math. Comp., 68 (1999), pp. 169–186.

[44] E. Wegert, *Visual complex functions. An introduction with phase portraits.*, Birkhäuser/Springer Basel AG, Basel, 2012.

[45] A. S. Wilmshurst, *The valence of harmonic polynomials*, Proc. Amer. Math. Soc., 126 (1998), pp. 2077–2081.

[46] E. Zeidler, *Nonlinear functional analysis and its applications. I: Fixed-point theorems*, Springer-Verlag, New York, 1986.

Number and location of pre-images under harmonic mappings in the plane*

Olivier Sète† Jan Zur†

Abstract

We derive a formula for the number of pre-images under a non-degenerate harmonic mapping f, using the argument principle. This formula reveals a connection between the pre-images and the caustics. Our results allow to deduce the number of pre-images under f geometrically for every non-caustic point. We approximately locate the pre-images of points near the caustics. Moreover, we apply our results to prove that for every $k = n, n+1, \dots, n^2$ there exists a harmonic polynomial of degree n with k zeros.

Keywords: Harmonic mappings, pre-images, caustics, argument principle, valence, zeros of harmonic polynomials.

Mathematics subject classification (2010): 30C55; 31A05; 55M25.

1 Introduction

Harmonic mappings in the plane, i.e., functions $f : \Omega \to \mathbb{C}$ with $\Delta f = 0$ on an open set $\Omega \subseteq \mathbb{C}$, regained attention in the last decades, starting from the seminal work of Clunie and Sheil-Small [10]. See, e.g., the large collection of open problems by Bshouty and Lyzzaik [9] and references therein. While we consider here multivalent harmonic mappings, also (locally) univalent harmonic mappings are of interest, see, e.g., Duren's textbook [11], especially in the context of quasi-conformal mappings [1].

Numerous authors have studied the number and location of zeros of harmonic mappings, i.e., the solutions of $f(z) = 0$. Of particular interest have been harmonic polynomials of the form $f(z) = p(z) - \overline{z}$ [19, 13],

*A final version of this preprint manuscript appeared as follows: O. Sète and J. Zur. Number and location of pre-images under harmonic mappings in the plane. *Ann. Fenn. Math.* (2021) 46(1), pp. 225–247, doi:10.5186/aasfm.2021.4614.

†TU Berlin, Institute of Mathematics, MA 3-3, Straße des 17. Juni 136, 10623 Berlin, Germany. {sete,zur}@math.tu-berlin.de

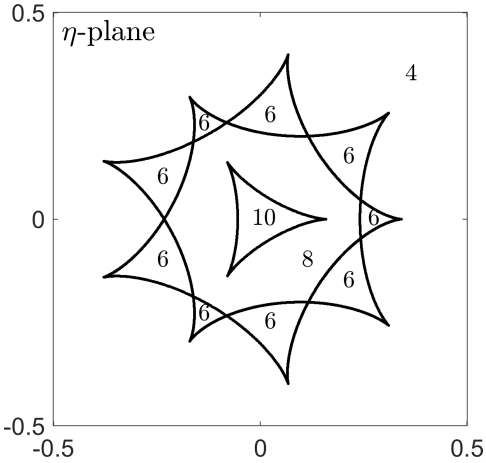


Figure 1: Number of pre-images of η under $f(z) = z - \overline{z^2/(z^3 - 0.6^3)}$ for an η in the respective regions; see also Example 3.9 below. The black lines mark the caustics (critical values) of f. The number of pre-images of η in the outer tile corresponds to the number of poles of f (including ∞).

or $f(z) = p(z) + \overline{q(z)}$ and the questions related to Wilmshurst's conjecture [38, 20, 16]. Also, the zeros of rational harmonic mappings of the form $f(z) = r(z) - \overline{z}$ have been studied intensively [17, 7, 25, 26, 22], since these are of interest when modeling the phenomenon of gravitational lensing [18, 29, 5].

Here we focus on solutions of $f(z) = \eta$ for given (but arbitrary) $\eta \in \mathbb{C}$. As shown in [21] for rational harmonic mappings of the form $f(z) = r(z) - \overline{z}$, the number of solutions can vary significantly under changes of η. Moreover, changes only occur when η is "moved" through the caustics of f; see Figure 1. This paper is devoted to study this effect for a more general class of harmonic mappings. We show the following:

(1) In Section 3 we derive (local and global) formulas for the number of pre-images of η under a non-degenerate harmonic mapping f (Definition 3.1) in terms of the poles and the winding number of the caustics about η, e.g.,

$$N_\eta(f) = P(f) + 2 \sum_{\gamma \in \mathrm{crit}} n(f \circ \gamma; \eta); \tag{1.1}$$

see Theorem 3.4. An immediate consequence of (1.1) is that the number of pre-images changes by ± 2 when η changes from one side to the other of a single caustic arc; see Figure 1.

(2) In Section 4 we complement Lyzzaik's study [27] of the local behavior of light harmonic mappings at their critical points. We approximately locate pre-images of η near a fold caustic point, which makes the pre-images also accessible for computations. Moreover, we determine for which η near a fold we have locally two or no pre-images; see Theorem 4.2.

(3) In Section 5 we apply the results from Sections 3 and 4 to harmonic polynomials. In particular we prove that for all $k \in \{n, n+1, \ldots, n^2\}$

there exists a harmonic polynomial $f(z) = p(z) + \overline{q(z)}$ with $\deg(p) = n$ and $\deg(q) < n$ with exactly k zeros, i.e., every number between the minimum and maximum can be attained; see Corollary 5.6. This generalizes a result of Bleher et al. [7, Thm. 1.1].

2 Preliminaries

The key ingredient to derive the formulas for the exact number of pre-images in Section 3 is the argument principle for harmonic mappings, applied on the critical set. In preparation, we collect and extend several known results in this section.

A *harmonic mapping* is a function $f : \Omega \to \mathbb{C}$ defined on an open set $\Omega \subseteq \mathbb{C}$ and with

$$\Delta f = \partial_{xx} f + \partial_{yy} f = 4\partial_{\overline{z}}\partial_z f = 0,$$

where ∂_z and $\partial_{\overline{z}}$ denote the *Wirtinger derivatives* of f; see e.g. [11, Sect. 1.2]. If f is harmonic in the open disk $D = \{z \in \mathbb{C} : |z - z_0| < r\}$, it has a local decomposition

$$f(z) = h(z) + \overline{g(z)} = \sum_{k=0}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=0}^{\infty} b_k (z - z_0)^k}, \quad z \in D, \tag{2.1}$$

with analytic functions h and g in D, which are unique up to an additive constant; see [12, p. 412] or [11, p. 7]. If f is harmonic in the punctured disk $D = \{z \in \mathbb{C} : 0 < |z - z_0| < r\}$, it has a local decomposition

$$f(z) = \sum_{k=-\infty}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=-\infty}^{\infty} b_k (z - z_0)^k} + c \log|z - z_0|, \quad z \in D; \tag{2.2}$$

see [35, 14]. We consistently use the notation from (2.1) and (2.2).

The *Jacobian* of a harmonic mapping f at $z \in \Omega$ is

$$J_f(z) = |\partial_z f(z)|^2 - |\partial_{\overline{z}} f(z)|^2 = |h'(z)|^2 - |g'(z)|^2, \tag{2.3}$$

where $f = h + \overline{g}$ is a local decomposition (2.1). We call f *sense-preserving* at z if $J_f(z) > 0$, *sense-reversing* at z if $J_f(z) < 0$, and *singular* at z if $J_f(z) = 0$. Moreover, we call f *singular*, if f is singular at one of its zeros. If φ is an analytic function, then $f \circ \varphi$ is again a harmonic mapping and

$$J_{f\circ\varphi}(z) = J_f(\varphi(z))|\varphi'(z)|^2. \tag{2.4}$$

In particular, if $\varphi'(z) \neq 0$, the maps f at $\varphi(z)$ and $f \circ \varphi$ at z are simultaneously sense-preserving, sense-reversing, or singular, respectively.

2.1 Critical set and caustics

The points at which a harmonic mapping f is singular form the *critical set*

$$\mathcal{C} = \{z \in \Omega : J_f(z) = 0\}, \tag{2.5}$$

which consists of the level set of an analytic function, and certain isolated points, as we see next.

The *second complex dilatation* of a harmonic mapping f is

$$\omega(z) = \frac{\overline{\partial_{\overline{z}} f(z)}}{\partial_z f(z)} = \frac{g'(z)}{h'(z)},$$

with the decomposition $f = h + \overline{g}$ from (2.1); see [11, p. 5], [1, p. 5] or [35, p. 71]. We assume that $\partial_z f = h'$ has only isolated zeros in Ω, so that ω is analytic in $\{z \in \Omega : \partial_z f(z) \neq 0\}$, and the singularities of ω in Ω are poles or removable singularities (which we assume to be removed). Moreover, we assume that $|\omega| \not\equiv 1$ on an open set (harmonic mappings with this property are characterized in [27, Lem. 2.1]).

Let $z_0 \in \Omega$. If $h'(z_0) \neq 0$, then $J_f(z_0) = |h'(z_0)|^2 - |g'(z_0)|^2 = 0$ is equivalent to $|\omega(z_0)| = 1$, and if $h'(z_0) = 0$, then $J_f(z_0) = 0$ is equivalent to $g'(z_0) = 0$. Hence, $|\omega(z_0)| = 1$ implies $J_f(z_0) = 0$, but the converse is not true in general. Define

$$\mathcal{M} = \{z \in \mathcal{C} : |\omega(z)| \neq 1\}. \tag{2.6}$$

By the above computation,

$$\mathcal{M} = \{z \in \Omega : h'(z) = g'(z) = 0 \text{ and } \lim_{\zeta \to z} |\omega(\zeta)| \neq 1\}.$$

For $z_0 \in \mathcal{M}$, there exists a neighborhood of z_0 containing no other point in $\mathcal{C}$; see [27, Lem. 2.2]. By construction,

$$\mathcal{C} \setminus \mathcal{M} = \{z \in \Omega : |\omega(z)| = 1\}$$

is a level set of the analytic function ω. Hence, $\mathcal{C} \setminus \mathcal{M}$ consists of analytic curves, which intersect in $z_0 \in \mathcal{C} \setminus \mathcal{M}$ if and only if $\omega'(z_0) = 0$. More precisely, if $\omega^{(k)}(z_0) = 0$ for $k = 1, \ldots, n-1$ and $\omega^{(n)}(z_0) \neq 0$, then $2n$ analytic arcs meet at z_0 with equispaced angles [36, p. 18]; see also Example 3.11.

At points $z \in \mathcal{C} \setminus \mathcal{M}$ with $\omega'(z) \neq 0$, the equation

$$\omega(\gamma(t)) = e^{it} \tag{2.7}$$

implicitly defines a local analytic parametrization $z = \gamma(t)$ of $\mathcal{C} \setminus \mathcal{M}$. We can write it locally as $\gamma(t) = \omega^{-1}(e^{it})$ with a continuous branch of ω^{-1}. The corresponding tangent vector at $z = \gamma(t)$ is

$$\gamma'(t) = i\frac{\omega(z)}{\omega'(z)}. \tag{2.8}$$

By construction f is sense-preserving to the left of γ, and sense-reversing to the right of γ.

The image of the critical set under a harmonic mapping f plays a decisive role for the number of pre-images. We call the set of critical values of f, i.e., $f(\mathcal{C})$, the set of *caustic points*, or simply the *caustics* of f. An $\eta \in \mathbb{C}$ has a pre-image under f on the critical set if, and only if, η is a caustic point.

The next lemma characterizes a tangent vector to the caustics and the curvature of the caustics; see [27, Lem. 2.3].

Lemma 2.1. *Let f be a harmonic mapping, $z_0 \in \mathcal{C} \setminus \mathcal{M}$ with $\omega'(z_0) \neq 0$, and let $z_0 = \gamma(t_0)$ with the parametrization (2.7). Then $f \circ \gamma$ is a parametrization of a caustic and the corresponding tangent vector at $f(z_0)$ is*

$$\tau(t_0) = \frac{d}{dt}(f \circ \gamma)(t_0) = e^{-it_0/2}\psi(t_0),$$

with

$$\psi(t_0) = 2\,\mathrm{Re}(e^{it_0/2}h'(\gamma(t_0))\gamma'(t_0)),$$

where $f = h + \overline{g}$ is a decomposition (2.1) in a neighborhood of z_0. In particular, the rate of change of the argument of the tangent vector is

$$\frac{d}{dt}\arg(\tau(t))\big|_{t=t_0} = -\frac{1}{2}$$

at points where $\psi(t_0) \neq 0$, i.e., the curvature of the caustics is constant with respect to the parametrization $f \circ \gamma$.

Moreover, ψ has either only finitely many zeros, or is identically zero, in which case f is constant on γ.

Definition 2.2. In the notation of Lemma 2.1, assume that the tangent $\tau(t_0)$ exists. Then, the point $(f \circ \gamma)(t_0)$ is called

1. a *fold caustic point* or simply a *fold*, if the tangent is non-zero,
2. a *cusp* of the caustic, if ψ has a zero with a sign change at t_0.

Remark 2.3.

1. If $(f \circ \gamma)(t_0)$ is a fold, then f is *light* (i.e., $f^{-1}(\{\eta\})$ is empty or totally disconnected for every $\eta \in \mathbb{C}$) in a neighborhood of $z_0 = \gamma(t_0)$. Indeed, if $\mathcal{C} \setminus \mathcal{M}$ can be parametrized according to (2.7), then J_f is not identically zero. Also, $f \circ \gamma$ is not constant near t_0. Hence, f is light in a neighborhood of z_0 by [27, Thm. 2.1].
2. At a cusp, the tangent vector becomes zero and the argument of the tangent vector jumps by $+\pi$. Note that the caustic either has only a finite number of cusps, or degenerates to a single point by Lemma 2.1.
3. In [27, Def. 2.2], a critical point $z_0 = \gamma(t_0)$ is called a critical point of (i) the first kind, if $f(z_0)$ is a cusp, (ii) the second kind, if $h'(z_0) = 0$ or $g'(z_0) = 0$, and if $\psi(t_0) = 0$ but ψ does not change its sign, and (iii) the third kind, if $\omega'(z_0) = 0$.

The curvature and the cusps of the caustics of f are apparent in the examples in Figure 4. The next lemma characterizes the fold caustic points in terms of the coefficients in (2.1).

Lemma 2.4. *Let f be a harmonic mapping, $z_0 \in \mathcal{C} \setminus \mathcal{M}$ with $\omega'(z_0) \neq 0$ and $h'(z_0) \neq 0$, and let $z_0 = \gamma(t_0)$ with the parametrization (2.7). We consider the decomposition (2.1) of f at z_0 and define $\theta \in [0, \pi[$ by $\overline{b}_1 = a_1 e^{i2\theta}$. Then the following are equivalent:*

1. $\psi(t_0) \neq 0$,
2. $\operatorname{Im}\left(\dfrac{1}{e^{it_0/2}a_1}\left(\dfrac{a_2}{a_1} - \dfrac{b_2}{b_1}\right)\right) \neq 0$,
3. $\operatorname{Im}\left(\dfrac{a_2}{a_1}e^{i\theta} + \overline{\left(\dfrac{b_2}{b_1}e^{i\theta}\right)}\right) \neq 0$.

Proof. Using (2.8), $e^{it_0} = \omega(z_0) = b_1/a_1$ and $\omega'(z_0) = 2\frac{b_2a_1 - b_1a_2}{a_1^2}$, we have

$$0 \neq \psi(t_0) = 2\operatorname{Re}\left(e^{it_0/2}h'(z_0)i\frac{\omega(z_0)}{\omega'(z_0)}\right) = \operatorname{Re}\left(ie^{it_0/2}a_1\frac{b_1a_1}{b_2a_1 - b_1a_2}\right).$$

Since $\operatorname{Re}(z) \neq 0$ if and only if $\operatorname{Re}(1/z) \neq 0$ (for $z \neq 0$), this is equivalent to

$$0 \neq \operatorname{Re}\left(-i\frac{1}{e^{it_0/2}a_1}\frac{b_2a_1 - b_1a_2}{b_1a_1}\right) = -\operatorname{Im}\left(\frac{1}{e^{it_0/2}a_1}\left(\frac{a_2}{a_1} - \frac{b_2}{b_1}\right)\right).$$

Write $a_1 = |a_1|e^{i\alpha}$, then $b_1 = a_1e^{it_0} = \overline{a}_1e^{-i2\theta}$ implies $e^{i(2\alpha+t_0)} = e^{-i2\theta}$, and hence $e^{it_0/2}a_1 = \pm|a_1|e^{-i\theta}$, which yields the equivalence of 2. and 3. □

2.2 The argument principle for harmonic mappings

Let f be continuous and non-zero on the trace of a curve $\gamma : [a, b] \to \mathbb{C}$. Then the *winding of f on γ* is defined as the change of argument of $f(z)$ as z travels along γ from $\gamma(a)$ to $\gamma(b)$, divided by 2π, i.e.,

$$W(f; \gamma) = \frac{1}{2\pi}\Delta_\gamma \arg(f(z)) = \frac{1}{2\pi}(\theta(b) - \theta(a)), \tag{2.9}$$

where $\theta : [a, b] \to \mathbb{R}$ is continuous with $\theta(t) = \arg(f(\gamma(t))$; see [3, Sect. 2.3] or [4, Ch. 7] for details.

Let now γ be a closed curve. We denote the *winding number* of γ about $\eta \in \mathbb{C} \setminus \operatorname{trace}(\gamma)$ by $n(\gamma; \eta)$, which is related to the winding through

$$W(f; \gamma) = n(f \circ \gamma; 0) \quad \text{and} \quad n(\gamma; \eta) = W(z \mapsto z - \eta; \gamma). \tag{2.10}$$

In particular, $W(f; \gamma)$ is an integer. Note that $W(f; \gamma) = n(f \circ \gamma; 0) = 0$ if f is constant on γ. Moreover, the winding is also called the degree or topological degree of f on γ; see [23, p. 3] or [34, p. 29].

The argument principle for a continuous function f relates the winding of f to the indices of its exceptional points. A point $z_0 \in \mathbb{C}$ is called an *isolated exceptional point* of a function f, if f is continuous and non-zero in a punctured neighborhood $D = \{z \in \mathbb{C} : 0 < |z - z_0| < r\}$ of z_0, and if f is either zero, not continuous, or not defined at z_0. Then the *Poincaré index* of f at z_0 is defined as

$$\operatorname{ind}(f; z_0) = W(f; \gamma), \tag{2.11}$$

where γ is a closed Jordan curve in D about z_0 oriented in the positive sense, i.e., with $n(\gamma; z_0) = 1$. The Poincaré index is also called the *index* [23, Def. 2.2.2] or the multiplicity [34, p. 44]. Similarly, ∞ is an isolated exceptional point of f, if f is continuous and non-zero in $D = \{z \in \mathbb{C} : |z| > R\}$. We define $\operatorname{ind}(f; \infty) = W(f; \gamma)$, where γ is a closed Jordan curve in D which is negatively oriented and surrounding the origin, such that ∞ lies on the left of γ on the Riemann sphere $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$. In either case the Poincaré index is independent of the choice of γ. We get with $\varphi(z) = z^{-1}$

$$\operatorname{ind}(f; \infty) = W(f; \gamma) = W(f \circ \varphi; \varphi^{-1} \circ \gamma) = \operatorname{ind}(f \circ \varphi; 0). \tag{2.12}$$

The Poincaré index generalizes the multiplicity of zeros and order of poles of an analytic function; see e.g. [34, p. 44].

The following version of the argument principle for continuous functions can be obtained from [3, Sect. 2.3], or [34, Sect. 2.3]. Special versions for harmonic mappings are given in [12] and [35, Thm. 2.2].

Theorem 2.5 (Argument principle)**.** *Let D be a multiply connected domain in $\widehat{\mathbb{C}}$ whose boundary consists of Jordan curves $\gamma_1, \ldots, \gamma_n$, which are oriented such that D is on the left. Let f be continuous and non-zero in $\overline{D}$, except for finitely many exceptional points $z_1, \ldots, z_k \in D$. We then have*

$$\sum_{j=1}^{n} W(f; \gamma_j) = \sum_{j=1}^{k} \operatorname{ind}(f; z_j).$$

Using the argument principle and the definition of the Poincaré index at infinity yields the following theorem.

Theorem 2.6. *Let f be defined, continuous and non-zero on $\widehat{\mathbb{C}}$, except for finitely many isolated exceptional points $z_1, \ldots, z_n$ in $\widehat{\mathbb{C}}$, then*

$$\sum_{j=1}^{n} \operatorname{ind}(f; z_j) = 0.$$

The exceptional points of a harmonic mapping f are its zeros and points where f is not defined. We determine their indices, beginning with the zeros; see [12, p. 413] or [35, p. 66].

Proposition 2.7. *Let f be a harmonic mapping with a zero z_0, such that the local decomposition* (2.1) *is of the form*

$$f(z) = \sum_{k=n}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=n}^{\infty} b_k (z - z_0)^k}, \quad n \geq 1,$$

where a_n or b_n can be zero, then

$$\operatorname{ind}(f; z_0) = \begin{cases} +n & \text{if } |a_n| > |b_n|, \\ -n & \text{if } |a_n| < |b_n|, \end{cases} \tag{2.13}$$

and, in particular,

$$\operatorname{ind}(f; z_0) = \begin{cases} +1 & \text{if } f \text{ is sense-preserving at } z_0, \\ -1 & \text{if } f \text{ is sense-reversing at } z_0. \end{cases} \tag{2.14}$$

A zero z_0 of a harmonic mapping f with $\operatorname{ind}(f; z_0) \in \mathbb{Z} \setminus \{-1, 1\}$ is a singular zero by the above result. Proposition 2.7 covers non-singular zeros and the zeros in $\mathcal{M}$; see (2.6). If $|a_n| = |b_n| \neq 0$, then z_0 is a singular zero in $\mathcal{C} \setminus \mathcal{M}$, in which case the determination of the index is more challenging; see [24] for the special case $f(z) = h(z) - \overline{z}$.

Remark 2.8. Zeros of f in $\mathcal{M}$ can be interpreted as multiple zeros of f. For a zero $z_0 \in \mathcal{M}$ of f, there exists $r > 0$ such that f is defined, non-zero and either sense-preserving or sense-reversing in $D = \{z \in \mathbb{C} : 0 < |z - z_0| \leq r\}$. For $0 < |\varepsilon| < m = \min_{|z-z_0|=r} |f(z)|$ and z with $|z - z_0| = r$ we have

$$|f(z) + \varepsilon - f(z)| = |\varepsilon| < m \leq |f(z)|,$$

which implies $W(f + \varepsilon; \gamma) = W(f; \gamma) = \operatorname{ind}(f; z_0)$ by Rouché's theorem; see e.g. [32, Thm. 2.3]. Since $f + \varepsilon$ has no poles in $\overline{D}$ and $f(z_0) + \varepsilon \neq 0$, it has $|\operatorname{ind}(f; z_0)|$ many distinct zeros in D by the argument principle.

Isolated exceptional points where f is not defined are classified according to the limit $\lim_{z \to z_0} f(z)$; see [35, Def. 2.1], [34, p. 44], and the classical notions for real-valued harmonic functions, e.g. [15, §15.3, III].

Definition 2.9. Let f be a harmonic mapping in a punctured disk around $z_0 \in \mathbb{C}$. Then z_0 is called

1. a *removable singularity* of f, if $\lim_{z \to z_0} f(z) = c \in \mathbb{C}$,
2. a *pole* of f, if $\lim_{z \to z_0} f(z) = \infty$,
3. an *essential singularity* of f, if $\lim_{z \to z_0} f(z)$ does not exist.

If one defines $f(z_0) = c$ at a removable singularity, then f is harmonic in z_0; apply [15, Thm. 15.3d] to the real and imaginary parts of f. In the sequel, we assume that removable singularities have been removed. If $c = 0$, then z_0 is a zero of f, and still an exceptional point.

For most poles of harmonic mappings, the Poincaré index can be determined from the decomposition (2.2).

Proposition 2.10. *Let f be a harmonic mapping in a punctured neighborhood of z_0, such that the local decomposition* (2.2) *is of the form*

$$f(z) = \sum_{k=-n}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=-n}^{\infty} b_k (z - z_0)^k} + c \log|z - z_0|,$$

where a_{-n} or b_{-n} can be zero, then

$$\operatorname{ind}(f; z_0) = \begin{cases} -n & \text{if } n \geq 1 \text{ and } |a_{-n}| > |b_{-n}|, \\ +n & \text{if } n \geq 1 \text{ and } |a_{-n}| < |b_{-n}|, \\ 0 & \text{if } n = 0 \text{ and } c \neq 0. \end{cases}$$

Moreover, in each case z_0 is a pole of f. In the first case, f is sense-preserving near z_0, and in the second it is sense-reversing near z_0. In the third case, z_0 is an accumulation point of the critical set of f.

Proof. See [35, Lem. 2.2, 2.3, 2.4] for the first two cases. In the third case, we have $\operatorname{ind}(f; z_0) = 0$ by [35, pp. 70–71]. Moreover, ω can be continued analytically to $z_0 \notin \Omega$ with $|\omega(z_0)| = \lim_{z \to z_0} |\omega(z)| = 1$, since $\partial_z f(z) = \frac{c}{2} \frac{1}{z - z_0} + \sum_{k=1}^{\infty} a_k k (z - z_0)^{k-1}$ and $\partial_{\overline{z}} f(z) = \frac{\overline{c}}{2} \frac{1}{z - z_0} + \sum_{k=1}^{\infty} b_k k (z - z_0)^{k-1}$. Hence z_0 is an accumulation point of the critical set of f by the maximum modulus principle for ω. □

Remark 2.11. If $n \geq 1$ and $|a_{-n}| = |b_{-n}| \neq 0$, we have that:

1. z_0 is an accumulation point of the critical set of f, as in the proof,
2. z_0 is a pole or an essential singularity of f, and both cases occur. Consider $f_1(z) = z^{-2} + z^{-1} + \overline{z}^{-2}$ and $f_2(z) = z^{-2} + z + \overline{z}^{-2}$, for which $z_0 = 0$ is an isolated exceptional point. The origin is a pole of f_1, since $\lim_{z \to 0} f_1(z) = \infty$, and $\operatorname{ind}(f_1; 0) = 0$; see [35, Ex. 2.6]. In contrast, $\lim_{z \to 0} f_2(z)$ does not exist (compare the limits on the real axis and the lines with $\operatorname{Re}(z^{-2}) = 0$), i.e., f_2 has an essential singularity at 0.

3 The number of pre-images

For non-degenerate harmonic mappings f, we derive explicit formulas for the number of pre-images of a non-caustic point η, in terms of the poles of f and of the winding number of the caustics of f about η. The proofs are based on the argument principle. Moreover, we deduce geometrically the number of pre-images from the caustics.

Definition 3.1. We call a harmonic mapping f *non-degenerate*, if the following conditions hold:

1. f is defined in $\widehat{\mathbb{C}}$ with the possible exception of finitely many poles,

2. at a pole $z_0 \in \mathbb{C}$ of f, the decomposition (2.2) has the form

$$f(z) = \sum_{k=-n}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=-n}^{\infty} b_k (z - z_0)^k} + c \log|z - z_0|, \qquad (3.1)$$

with $n \geq 1$ and $|a_{-n}| \neq |b_{-n}|$. And if ∞ is a pole of f, then

$$f(z) = \sum_{k=-\infty}^{n} a_k z^k + \overline{\sum_{k=-\infty}^{n} b_k z^k} + c \log|z|, \quad \text{for } |z| > R, \qquad (3.2)$$

with $n \geq 1$ and $|a_n| \neq |b_n|$, and $R > 0$,

3. the critical set $\mathcal{C}$ of f is bounded.

Remark 3.2.

1. The first item in Definition 3.1 allows to apply the argument principle globally. By the second item, we can determine the Poincaré index of a pole with Proposition 2.10, and the poles are not accumulation points of $\mathcal{C}$; see Remark 2.11. In particular, $\mathcal{C}$ is a closed subset of $\mathbb{C}$.

2. Harmonic polynomials $f(z) = p(z) + \overline{q(z)}$ with $\deg(p) > \deg(q)$, and rational harmonic mappings $f(z) = r(z) - \overline{z}$ with $\lim_{z\to\infty} f(z) = \infty$ are non-degenerate. For these functions, the number of zeros or pre-images is intensively studied; see e.g. [38, 19, 17, 13, 7, 25, 26, 32, 20, 21, 22, 5, 16].

3. We discuss the difference between non-degenerate harmonic mappings and the maps in [27, 28]. By [27, Thm. 2.1], a harmonic mapping is either (a) light, (b) has a zero Jacobian, or (c) is constant on an analytic subarc of $\mathcal{C} \setminus \mathcal{M}$. While Lyzzaik [27] and Neumann [28] consider harmonic mappings that are light (case (a)) and have no poles, we allow cases (a) and (c) and certain poles. For example, the harmonic mapping $f(z) = \frac{1}{z} - \overline{z}$, modeling the Chang-Refsdal lens in gravitational lensing [2], is non-degenerate with poles at 0 and ∞, and with critical set $\mathcal{C} = \{z \in \mathbb{C} : |z| = 1\}$. It is not light, since $f(\mathcal{C}) = \{0\}$.

4. It is possible that different arcs of the critical set are mapped onto the same caustic arc; see Example 5.1.

3.1 A formula for the number of pre-images

To count the number of pre-images under f with the argument principle, we separate the regions where f is sense-preserving and sense-reversing.

Let f be a non-degenerate harmonic mapping. In particular, the critical set $\mathcal{C}$ is bounded and closed. For each connected component Γ of $\mathcal{C} \setminus \mathcal{M}$, we construct a single closed curve γ parametrizing Γ and traveling through every critical arc exactly once, according to (2.7). There are two possibilities.

1. If ω' is non-zero on Γ, then Γ is the trace of a closed Jordan curve γ.

2. If ω' has zeros on Γ, then Γ consists of Jordan arcs that meet at the zeros of ω', and we proceed as follows. We interpret the component Γ as a directed multigraph with intersection points as vertices and critical arcs as arcs of the graph, directed in the sense of (2.7). At a vertex corresponding to an $(n-1)$-fold zero of ω', $2n$ arcs meet. Due to the orientation of the arcs, the same number of arcs are incoming and outgoing. Hence we find an Euler circuit in the graph [8, Sect. I.3], which corresponds to the desired parametrization γ of Γ.

We call the above γ a *critical curve*, and denote the set of all these curves by crit; see Figure 4 below for examples.

The critical set induces a partition of $\widehat{\mathbb{C}} \setminus \mathcal{C}$ into open and connected components A, where $\partial A \subseteq \mathcal{C}$ and f is either sense-preserving or sense-reversing on A (more precisely on A minus the poles of f). Such a component may or may not be simply connected; see Figure 4 (top left). Denote the component containing ∞ by A_∞. For $A \neq A_\infty$, note that ω has at least one zero/pole in A if f is sense-preserving/sense-reversing in A, by the minimum modulus principle/maximum modulus principle for ω. If ω is identically zero/infinity, then f is analytic/anti-analytic, and there is only one component. Otherwise, ω has only finitely many zeros and poles on the compact set $\widehat{\mathbb{C}} \setminus A_\infty$, and there are at most finitely many other components, and we write

$$\mathcal{A} = \{A_1, \ldots, A_m\}. \tag{3.3}$$

This generalizes a similar partition for rational harmonic mappings of the form $f(z) = r(z) - \overline{z}$ from [21, Sect. 2].

For $A \in \mathcal{A}$, we construct parametrizations $\gamma_1, \ldots, \gamma_n$ according to (2.7) of the connected components $\Gamma_1, \ldots, \Gamma_n$ of $\Gamma = (\partial A) \setminus \mathcal{M}$. If ω' is non-zero on Γ_j, then there exists a closed Jordan curve γ_j with $\operatorname{trace}(\gamma_j) = \Gamma_j$ as before. Otherwise we interpret Γ_j as a directed multigraph and show the existence of an Euler circuit as above. For a zero $z_0 \in \Gamma_j$ of ω' the set $A_\varepsilon = \{z \in A : 0 < |z - z_0| < \varepsilon\}$ consists of k connected components for $\varepsilon > 0$ sufficiently small. Every component of A_ε produces one ingoing and one outgoing arc at the vertex corresponding to z_0; see Figure 2 (left). Hence, there exists an Euler circuit in Γ_j and we denote by γ_j a parametrization according to (2.7) of this circuit. Applying the above construction to all $A \in \mathcal{A}$ yields not necessarily a disjoint partition of $\mathcal{C} \setminus \mathcal{M}$, see Figure 4 (bottom left), and hence cannot be used in Theorem 3.4. In particular γ_j is potentially not a critical curve.

We determine the number of pre-images in one component $A \in \mathcal{A}$.

Theorem 3.3. *Let f be a non-degenerate harmonic mapping, $A \in \mathcal{A}$, and let $\gamma_1, \ldots, \gamma_n$ be a parametrization of $\Gamma = (\partial A) \setminus \mathcal{M}$ as above. Moreover,*

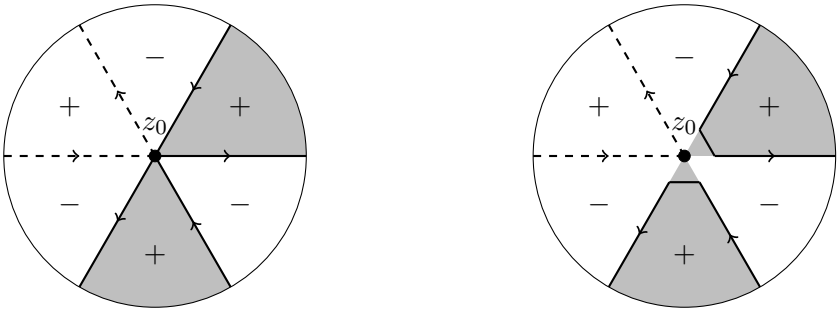


Figure 2: Left: A_ε (shaded) and oriented critical arcs near a zero z_0 of ω'. Right: Deformation of γ_j in the proof of Theorem 3.3. The $+/-$ signs indicate regions where f is sense-preserving/sense-reversing.

let $z_1, \dots, z_k$ be the poles of f in A, and define $P(f;A) = \sum_{j=1}^{k} |\mathrm{ind}(f; z_j)|$. Then, for $\eta \in \mathbb{C} \setminus f(\partial A)$, the number $N_\eta(f;A)$ of pre-images of η under f in A is

$$N_\eta(f;A) = P(f;A) + \sum_{j=1}^{n} n(f \circ \gamma_j; \eta). \tag{3.4}$$

Proof. We apply the argument principle to $f_\eta = f - \eta$ on A. Note that f_η is also non-degenerate, $J_f = J_{f_\eta}$, and f_η has the same poles with same index as f, so that $P(f_\eta; A) = P(f; A)$. Since f_η is non-zero on ∂A, it has no zeros in $\mathcal{M} \cap \overline{A}$. Moreover, f_η has only finitely many zeros in A. For a bounded A this holds since non-singular zeros are isolated [12, p. 413]. For A_∞, assume that f_η has infinitely many zeros in A_∞ and hence in some $\{z \in \mathbb{C} : |z| \geq R\}$. Then $f_\eta(1/z)$ has infinitely many non-singular zeros in $\{z \in \mathbb{C} : |z| \leq 1/R\}$, which contradicts the fact that such zeros are isolated.

First, suppose that Γ is non-empty and that $\gamma_1, \dots, \gamma_n$ are closed Jordan curves. If f_η is sense-preserving in A, then A lies to the left of $\gamma_1, \dots, \gamma_n$. The argument principle implies

$$\sum_{j=1}^{n} W(f_\eta; \gamma_j) = N_0(f_\eta; A) + \sum_{j=1}^{k} \mathrm{ind}(f_\eta; z_j) = N_\eta(f; A) - P(f; A),$$

where we used that f_η is sense-preserving and hence the index at a zero is $+1$ by (2.14) and negative at a pole by Proposition 2.10. We obtain (3.4) in this case with $W(f_\eta; \gamma_j) = n(f \circ \gamma_j; \eta)$; see (2.10). Recall that $n(f \circ \gamma_j; \eta) = 0$ if $f \circ \gamma_j$ is constant. If f_η is sense-reversing in A, then A lies to the right of $\gamma_1, \dots, \gamma_n$, and the index of f at a zero is -1 by (2.14) and positive at a pole by Proposition 2.10, and hence

$$\sum_{j=1}^{n} W(f_\eta; -\gamma_j) = -N_\eta(f; A) + P(f; A),$$

where $-\gamma_j$ denote the reversed curves. Since $W(f;-\gamma_j) = -W(f;\gamma_j)$, we obtain (3.4).

If some γ_j is not a Jordan curve, then it self-intersects at a zero z_0 of ω', as indicated in Figure 2 (left). However, f_η is continuous and non-zero at z_0. Hence, by an arbitrary small manipulation of γ_j, we obtain a Jordan curve on which f_η has the same winding. This is illustrated in Figure 2 (right). The proof then remains unchanged with the new curves.

Finally, if Γ is empty, then A is the only component in $\mathcal{A}$ and $N_\eta(f) = P(f)$ follows from Theorem 2.6. □

Summing over all $A \in \mathcal{A}$ gives the total number of pre-images.

Theorem 3.4. *Let f be a non-degenerate harmonic mapping. Then $N_\eta(f)$, the number of pre-images in $\widehat{\mathbb{C}}$ of $\eta \in \mathbb{C} \setminus f(\mathcal{C})$ under f, is*

$$N_\eta(f) = P(f) + 2 \sum_{\gamma \in \mathrm{crit}} n(f \circ \gamma; \eta).$$

Here $P(f) = \sum_{A \in \mathcal{A}} P(f;A)$ denotes the number of poles of f in $\widehat{\mathbb{C}}$ counted with the absolute values of their Poincaré indices, as in Theorem 3.3.

Proof. The function $f_\eta = f - \eta$ has no zeros on $\mathcal{C}$, since η is not a caustic point. Let $\mathcal{A} = \{A_1, \dots, A_m\}$ and denote by $\gamma_{1,j}, \dots, \gamma_{n_j,j}$ a parametrization of $(\partial A_j) \setminus \mathcal{M}$ as above. Applying Theorem 3.3 for $A_1, \dots, A_m$ yields

$$\begin{aligned} N_\eta(f) &= \sum_{j=1}^{m} N_\eta(f;A_j) = \sum_{j=1}^{m} \left(P(f;A_j) + \sum_{k=1}^{n_j} n(f \circ \gamma_{k,j}; \eta) \right) \\ &= P(f) + 2 \sum_{\gamma \in \mathrm{crit}} n(f \circ \gamma; \eta). \end{aligned}$$

Here we used that every $\gamma_{k,j}$ consists of arcs which are boundary arcs of exactly two components in $\mathcal{A}$, and that the critical curves are a (disjoint) parametrization of $\mathcal{C} \setminus \mathcal{M}$ according to (2.7). □

Remark 3.5. Theorems 3.3 and 3.4 not only contain a formula for counting the pre-images of η, but also allow to determine how the number of pre-images changes if η changes its position relative to the caustics of f. More precisely, the number of pre-images in $A \in \mathcal{A}$ changes by ± 1 if η "crosses" a single caustic arc from $f(\partial A)$; see Theorem 3.3.

For large enough $|\eta|$, the pre-images are near the poles. This generalizes [21, Thm. 3.1]. We write $D_\varepsilon(z_0) = \{z \in \mathbb{C} : |z - z_0| < \varepsilon\}$.

Theorem 3.6. *Let f be a non-degenerate harmonic mapping with poles $z_1, \dots, z_n$, let $\varepsilon > 0$ be such that the sets $D_\infty = \{z \in \mathbb{C} : |z| > \varepsilon^{-1}\}$ and $D_\varepsilon(z_1), \dots, D_\varepsilon(z_n)$ are disjoint, and such that on each set f is either*

sense-preserving or sense-reversing. Then, for every $\eta \in \mathbb{C}$ with $|\eta|$ large enough, we have

$$N_\eta(f; D_\varepsilon(z_k)) = |\mathrm{ind}(f; z_k)| \quad \textit{and} \quad N_\eta(f; D_\infty) = |\mathrm{ind}(f - \eta; \infty)|.$$

Moreover, all pre-images of η are in $D = \cup_{k=1}^n D_\varepsilon(z_k) \cup D_\infty$.

Proof. Let $\eta \in \mathbb{C}$ be such that $|f(z)| < |\eta|$ for $z \in \partial D$, which is possible since ∂D is compact and f continuous. To apply Rouché's theorem (e.g. [32, Thm. 2.3]) to $f_\eta = f - \eta$ and $g(z) = -\eta$, note that

$$|f_\eta(z) - g(z)| = |f(z)| < |\eta| \quad \text{for } z \in \partial D.$$

Since f is either sense-preserving or sense-reversing on $D_\varepsilon(z_k)$, we have

$$0 = W(g; \gamma_k) = W(f_\eta; \gamma_k) = \pm N_\eta(f; D_\varepsilon(z_k)) + \mathrm{ind}(f; z_k),$$

with $\gamma_k : [0, 2\pi] \to \mathbb{C}$, $\gamma_k(t) = z_k + \varepsilon e^{it}$. Hence, $N_\eta(f; D_\varepsilon(z_k)) = |\mathrm{ind}(f; z_k)|$ as in Theorem 3.3. Similarly, let $\gamma_\infty : [0, 2\pi] \to \mathbb{C}$, $\gamma_\infty(t) = \varepsilon^{-1} e^{-it}$, then

$$0 = W(g; \gamma_\infty) = W(f_\eta; \gamma_\infty) = \pm N_\eta(f; D_\infty) + \mathrm{ind}(f_\eta; \infty).$$

By increasing $|\eta|$, so that η lies outside all caustics, i.e., $n(f \circ \gamma; \eta) = 0$ for all $\gamma \in \mathrm{crit}$, we have with Theorem 3.4

$$N_\eta(f) = P(f) = \sum_{j=1}^n N_\eta(f; D_{\varepsilon_j}(z_j)) + N_\eta(f; D_\infty).$$

This implies that all pre-images of η are in D. □

Note that the number of pre-images determined in Theorem 3.6 is not necessarily the minimal number of pre-images as η ranges over $\mathbb{C} \setminus f(\mathcal{C})$; see Example 3.10 and Figure 4. For non-singular harmonic polynomials, however, this is the lower bound for the number of zeros; see the discussion at the beginning of Section 5.

We now consider η as *variable parameter*, and deduce the number of pre-images of η_2 from the number of pre-images of another point η_1, e.g., with sufficiently large $|\eta_1|$ as in Theorem 3.6.

The caustics induce a partition of $\mathbb{C} \setminus f(\mathcal{C})$ into open and connected components, which we call *caustic tiles*. This partition does not coincide with $f(\mathcal{A})$ in general, since f has not the open mapping property; see also Figure 4, where $\widehat{\mathbb{C}} \setminus \mathcal{C}$ and $\mathbb{C} \setminus f(\mathcal{C})$ have a different number of (connected) components. The winding number of $f \circ \gamma$ about η depends on the position of η with respect to the caustics, i.e., to which caustic tile η belongs to. The next theorem is an immediate and very useful consequence of Theorem 3.4.

Theorem 3.7. *For a non-degenerate harmonic mapping f and non-caustic points $\eta_1, \eta_2 \in \mathbb{C} \setminus f(\mathcal{C})$, we have*

$$N_{\eta_2}(f) = N_{\eta_1}(f) + 2 \sum_{\gamma \in \text{crit}} \big(n(f \circ \gamma; \eta_2) - n(f \circ \gamma; \eta_1) \big), \tag{3.5}$$

and in particular:

1. *If η_1 and η_2 are in the same caustic tile, then the number of pre-images under f is the same, i.e., $N_{\eta_2}(f) = N_{\eta_1}(f)$.*
2. *If η_1 and η_2 are separated by a single caustic $f \circ \gamma$, then the number of pre-images under f changes by two, i.e., $N_{\eta_2}(f) = N_{\eta_1}(f) \pm 2$.*
3. *$N_{\eta_1}(f)$ is odd if, and only if, $N_{\eta_2}(f)$ is odd.*
4. *Let $\eta_1, \eta_2 \in \mathbb{C}$. If $N_{\eta_1}(f)$ is even and $N_{\eta_2}(f)$ is odd, then η_1 or η_2 is a caustic point of f.*

We obtain a formula similar to (3.5) for each set $A \in \mathcal{A}$, using Theorem 3.3 instead of Theorem 3.4. This yields $N_{\eta_2}(f; A) = N_{\eta_1}(f; A)$ in 1. In 2., the number of pre-images increases/decreases by 1 in the sets A adjacent to the critical arc γ, and stays the same in all other sets A.

Items 3 and 4 are in the spirit of the "odd number of images theorem" from the theory of gravitational lensing in astrophysics [30, Thm. 11.5].

3.2 Counting pre-images geometrically

We determine geometrically whether the number of pre-images increases or decreases in item 2 of Theorem 3.7. The key ingredient is the curvature of the caustics (Lemma 2.1), which allows to spot their orientation in a plot; see Figure 3. Then, the change of the winding number $n(f \circ \gamma; \eta_2) - n(f \circ \gamma; \eta_1)$ can be determined with the next result.

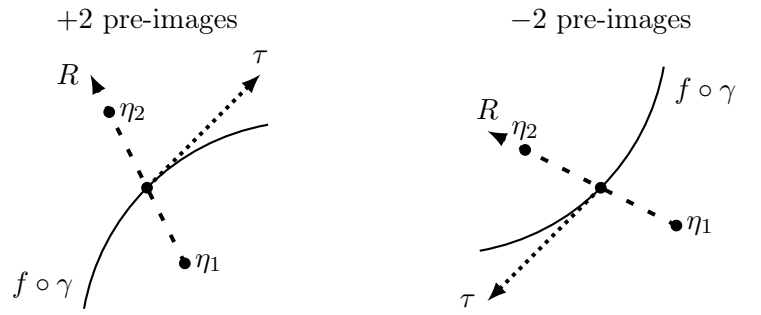
Proposition 3.8 ([31, Prop. 3.4.4])**.** *Let γ be a smooth closed curve and $\eta \notin \text{trace}(\gamma)$. Let further R be a ray from η to ∞ in direction $e^{i\varphi}$, such that R is not a tangent at any point on γ. Then R intersects γ at finitely many points $\gamma(t_1), \dots, \gamma(t_k)$ and we have for the winding number of γ about η*

$$n(\gamma; \eta) = \sum_{j=1}^{k} i_{t_j}(\gamma; R),$$

where the intersection index *i_{t_j} of γ and R at $\gamma(t_j)$, is defined by*

$$i_{t_j}(\gamma; R) = \begin{cases} +1, & \text{if } \operatorname{Im}(e^{-i\varphi}\gamma'(t_j)) > 0, \\ -1, & \text{if } \operatorname{Im}(e^{-i\varphi}\gamma'(t_j)) < 0. \end{cases}$$

Recall that $e^{i\varphi}$ and $\gamma'(t_j)$ form a right-handed basis if $\operatorname{Im}(e^{-i\varphi}\gamma'(t_j)) > 0$, and a left-handed basis if the imaginary part is negative.



$n(f \circ \gamma; \eta_2) = n(f \circ \gamma; \eta_1) + 1$ $\qquad$ $n(f \circ \gamma; \eta_2) = n(f \circ \gamma; \eta_1) - 1$

Figure 3: Intersection index and caustics in the η-plane.

Let η_1, η_2 be in two adjacent caustic tiles separated by a single caustic arc. We call two sets adjacent, if they share a common boundary arc. Consider the ray R from η_1 to ∞ through η_2, and let it intersect the caustic between η_1 and η_2 at a fold point $(f \circ \gamma)(t_0)$. Although the caustics are only piecewise smooth, we can smooth the finitely many (see Lemma 2.1) cusps as in [16, p. 16] to obtain a smooth curve with same winding numbers about η_1 and η_2. Then $n(f \circ \gamma; \eta_1) = n(f \circ \gamma; \eta_2) + i_{t_0}(f \circ \gamma; R)$ by Proposition 3.8, and equivalently

$$n(f \circ \gamma; \eta_2) - n(f \circ \gamma; \eta_1) = -i_{t_0}(f \circ \gamma; R),$$

where the intersection index is $+1$ if $\eta_2 - \eta_1$ and $\tau(t_0)$ form a right-handed basis, and -1 if the two vectors form a left-handed basis; see Figure 3.

Caustic tiles have three different shapes. We call a caustic tile B *deltoid-like* (respectively *cardioid-like*), if for every point $z_0 \in \partial B$, for which the tangent to the caustics exists and is non-zero, there exists an open disk D centered at z_0 such that the intersection of D and the tangent line to ∂B at z_0 is contained in B (respectively contained in $\mathbb{C} \setminus B$). We call a caustic tile *mixed*, if it is neither deltoid nor cardioid-like. In Figure 4 (middle right), the tiles with the number 6 are deltoid-like, the tile with the number 2 is cardioid-like, and the tile with the number 4 is a mixed caustic tile. Entering a deltoid-like tile gives two additional pre-images, entering a cardioid-like tile gives two fewer pre-images, for a mixed tile both occur according to the shape of the "crossed" caustic arc; see Figure 3 and Example 3.10.

Example 3.9. Consider the non-degenerate rational harmonic mapping

$$f(z) = z - \overline{\left(\frac{z^2}{z^3 - 0.6^3}\right)}.$$

Figure 4 (top) shows the critical set and the caustics of f. We have $P(f) = 4$, since f has four simple poles (∞ with index -1, the others with index 1); see Proposition 2.10. Thus, for η in the outer region, i.e., with $n(f \circ \gamma_j; \eta) = 0$

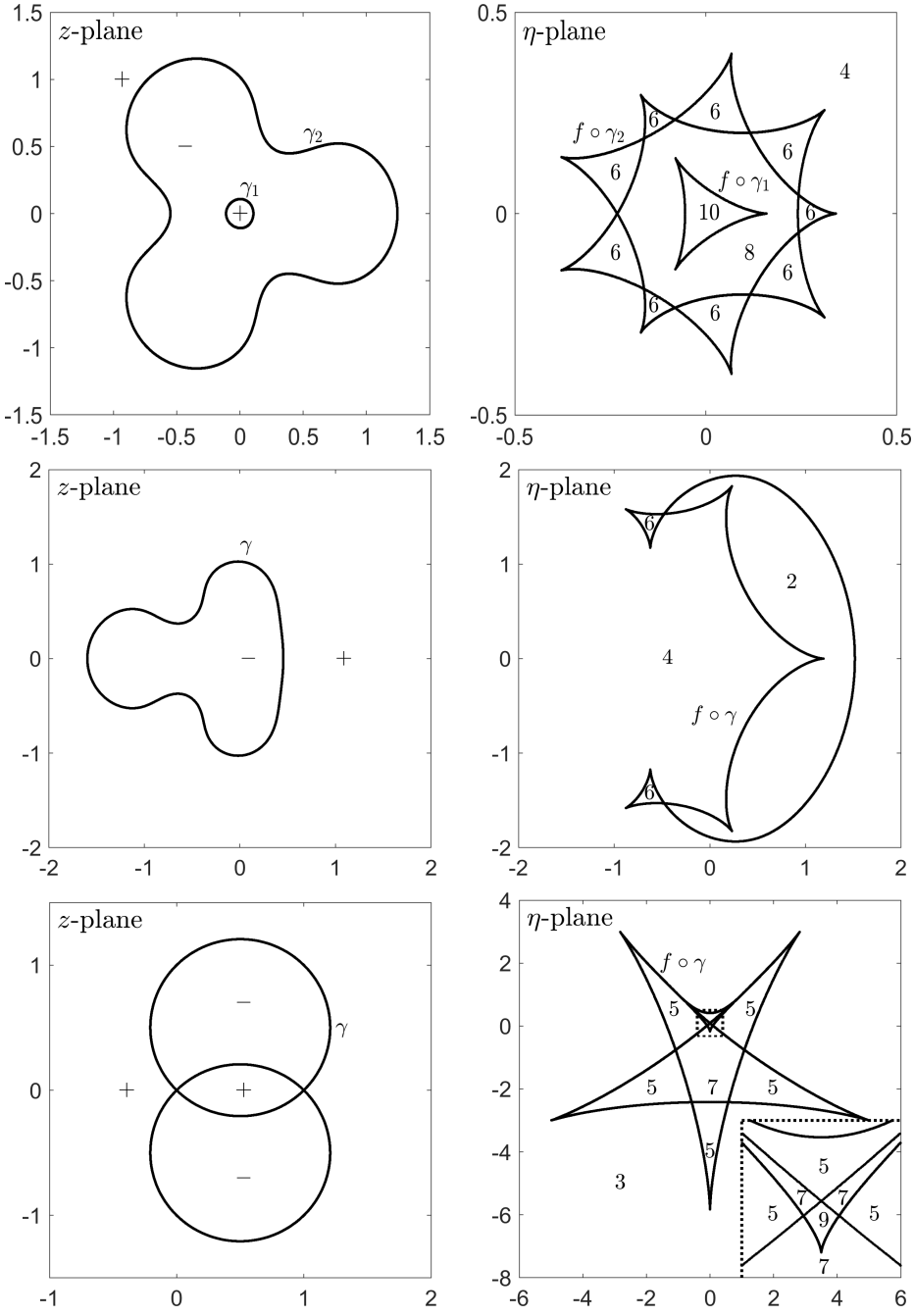


Figure 4: Critical curves (left) and caustics (right) of the functions in Examples 3.9 (top), 3.10 (middle), and 3.11 (bottom). The +/− signs indicate the regions where f is sense-preserving/sense-reversing. The numbers indicate the number of pre-images of an η in the respective caustic tile. The dotted line in the bottom right plot marks a zoom-in.

for $j = 1, 2$, we have $N_\eta(f) = 4+2\cdot 0 = 4$. For $\eta = 0$, we have $n(f\circ\gamma_1; 0) = 1$ and $n(f \circ \gamma_2; 0) = 2$, so that f has $N_0(f) = 4 + 2 \cdot 3 = 10$ zeros.

Certain rational harmonic mappings are studied in gravitational lensing in astrophysics; see e.g. [18, 25]. Also transcendental functions such as $f(z) = z - \overline{k/\sin(z)}$ appear in this context [6].

Example 3.10. Figure 4 (middle) shows the critical curves and caustics of the non-degenerate harmonic mapping

$$f(z) = z^2 + \frac{1}{\overline{z}} + \frac{1}{\overline{z}+1} + 2\log|z|.$$

We have $P(f) = 4$ from the simple poles at 0 and -1 with index $+1$ and the double pole at ∞ with index -2; see Proposition 2.10. Consequently, any η in the outer region (i.e., with $n(f \circ \gamma; \eta) = 0$) has 4 pre-images. Note the effect of deltoid-like, cardioid-like and mixed caustic tiles described above: the tiles where η has 6 pre-images are deltoid-like, the tile where η has 2 pre-images is cardioid-like, and the outer tile is mixed.

Example 3.11. The non-degenerate harmonic polynomial

$$f(z) = p(z) + \overline{q(z)} = z^n + (z-1)^n + \overline{i(z-1)^n - iz^n}, \quad n \geq 1,$$

has the maximum number of n^2 zeros [38, p. 2080]. Its critical set consists of $n-1$ circles, intersecting in 0 and 1, and can be parametrized as discussed in Section 3.1. Figure 4 (bottom) shows the critical set and caustics for $n = 3$.

4 Location of pre-images near the critical set

In Section 3, we omitted the case when η is on a caustic. Here, we study the local effect when η "crosses" a caustic, i.e., when the number of pre-images changes. Since this is a local effect, the harmonic mappings are neither required to be globally defined nor to be non-degenerate.

Non-singular pre-images persist under a small change of η, which is an immediate consequence of the inverse function theorem.

Proposition 4.1. *Let f be a harmonic mapping defined in the open set $\Omega \subseteq \mathbb{C}$ and let f be non-singular at $z_0 \in \Omega$. Then there exist open neighborhoods $U \subseteq \Omega \setminus \mathcal{C}$ of z_0 and V of $f(z_0)$ such that each $\eta \in V$ has exactly one pre-image under f in U.*

Lyzzaik [27] investigated the local behavior of light harmonic mappings, defined on an open and simply connected subset of $\mathbb{C}$. His analysis relies upon the local transformation of f near a critical point $z_0 \in \mathcal{C}$ into standard mappings $h_2 \circ f \circ h_1^{-1}(z) = z^n$ or $h_2 \circ f \circ h_1^{-1}(z) = \overline{z}^n$, where h_1 and h_2 are sense-preserving homeomorphisms; see [27, Sect. 3] for details. If such

a standard mapping exists we write $f_{z_0} \sim z^n$ and $f_{z_0} \sim \overline{z}^n$ respectively. One of Lyzzaik's results is the following: Let $f(z_0)$ be a fold and U be a neighborhood of z_0. Then there exists a partition U_1, U_2 of $U \setminus \mathcal{C}$ with $f_{z_0} \sim z$ in $\overline{U}_1$ and $f_{z_0} \sim \overline{z}$ in $\overline{U}_2$. Similarly, if $f(z_0)$ is a cusp and $h'(z_0) \neq 0$, we have $f_{z_0} \sim z^3$ in $\overline{U}_1$, $f_{z_0} \sim \overline{z}$ in $\overline{U}_2$ or $f_{z_0} \sim z$ in $\overline{U}_1$, $f_{z_0} \sim \overline{z}^3$ in $\overline{U}_2$; see [27, Thm. 5.1]. This allows to determine the *valence*

$$V(f;U) = \sup_{\eta \in \mathbb{C}} N_\eta(f;U) = \sup_{\eta \in \mathbb{C}} |\{z \in U : f(z) = \eta\}|$$

of f in U. In particular we have

$$V(f; D_\varepsilon(z_0)) = \begin{cases} 2, & \text{if } f(z_0) \text{ is a fold,} \\ 3, & \text{if } f(z_0) \text{ is a cusp with } h'(z_0) \neq 0, \end{cases} \tag{4.1}$$

for sufficiently small $\varepsilon > 0$; see [27, Thm 5.1]. However, the above transformations are not immediately available for practical computations in general.

We complement Lyzzaik's work by investigating which values near a fold $\eta = f(z_0)$ have actually 2, 1 or no pre-images under f in $D_\varepsilon(z_0)$, and by approximately locating the pre-images for certain η. For this we use convergence results on the harmonic Newton iteration

$$z_{k+1} = z_k - \frac{\overline{h'(z_k)} f(z_k) - \overline{g'(z_k) f(z_k)}}{J_f(z_k)}, \quad k \geq 0, \tag{4.2}$$

from [33]. If the sequence (4.2) converges and all iterates z_k are in $D \subseteq \mathbb{C}$, then there exists a zero of f in $\overline{D}$. The proof of the next theorem relies on this strategy.

Theorem 4.2. *Let f be a light harmonic mapping and $z_0 \in \mathcal{C} \setminus \mathcal{M}$, such that $\eta = f(z_0)$ is a fold. Moreover, let*

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=0}^{\infty} b_k (z - z_0)^k} \quad \text{and} \quad c = -\left(\frac{a_2 \overline{b}_1}{a_1} + \frac{\overline{b}_2 a_1}{\overline{b}_1} \right).$$

Then, for each sufficiently small $\varepsilon > 0$, there exists a $\delta > 0$, such that for all $0 < t < \delta$ we have:

1. *$\eta + tc$ has exactly two pre-images under f in $D_\varepsilon(z_0)$,*
2. *η has exactly one pre-image under f in $D_\varepsilon(z_0)$,*
3. *$\eta - tc$ has no pre-image under f in $D_\varepsilon(z_0)$.*

In case 1, each disk $\{z \in \mathbb{C} : |z - z_\pm| \leq const \cdot t\}$, where $z_\pm = z_0 \pm i\sqrt{t\, \overline{b}_1 / a_1}$, contains one of the two pre-images, and f is sense-preserving at one and sense-reversing at the other.

Proof. Since $z_0 \in \mathcal{C}$ and $f(z_0)$ is a fold, we have $h'(z_0) \neq 0$, and hence $|g'(z_0)| = |h'(z_0)| \neq 0$. Then there exists $\theta \in [0, \pi[$ with $\overline{b}_1 = a_1 e^{i2\theta}$, and

$$c = -a_1 e^{i\theta} \left(\frac{a_2}{a_1} e^{i\theta} + \overline{\frac{b_2}{b_1} e^{i\theta}} \right)$$

is non-zero by Lemma 2.4.

1. We apply the harmonic Newton iteration (4.2) to the shifted function $f_{\eta+tc} = f - (\eta + tc)$ with initial points $z_\pm$. By [33, Lem. 5.1, Thm. 5.2] and their proofs, the respective sequences of iterates remain in $D_\pm$, and converge to two distinct zeros of $f_{\eta+tc}$ for all sufficiently small $t > 0$. Thus, $\eta + tc$ has exactly two pre-images under f in $D_\varepsilon(z_0)$, using (4.1).

2. Since f is light and $f(z_0) = \eta$, there exists $\varepsilon > 0$ such that z_0 is the only pre-image of η in $D_\varepsilon(z_0)$.

3. We show first that the "direction" c is not tangential to the caustic, and hence that $\eta + tc$ and $\eta - tc$ are not in the same caustic tile. Since $\eta = f(z_0)$ is a fold, we have with $z_0 = \gamma(t_0)$ and the tangent τ from Lemma 2.1

$$\overline{\tau(t_0)} c = -\psi(t_0) e^{it_0/2} a_1 e^{i\theta} \left(\frac{a_2}{a_1} e^{i\theta} + \overline{\frac{b_2}{b_1} e^{i\theta}} \right) = \mp \psi(t_0) |a_1| \left(\frac{a_2}{a_1} e^{i\theta} + \overline{\frac{b_2}{b_1} e^{i\theta}} \right),$$

since $e^{it_0/2} e^{i\theta} a_1 = \pm |a_1|$; see the proof of Lemma 2.4. Since ψ is real, and non-zero at a fold, we have $\operatorname{Im}(\overline{\tau(t_0)} c) \neq 0$ by Lemma 2.4. Hence, for a sufficiently small $t > 0$, the points $\eta + tc$ and $\eta - tc$ are on different sides of the caustic $f \circ \gamma$, where γ denotes the critical curve through z_0. Thus, there are either $2 + 2 = 4$ or $2 - 2 = 0$ pre-images of $\eta - tc$ under f in $D_\varepsilon(z_0)$; see Theorem 3.4 if f is non-degenerate, and [28, Thm. 6.7] for light harmonic mappings. Since $V(f; D_\varepsilon(z_0)) = 2$ by (4.1), only the latter case is possible.

Moreover, the two pre-images of $\eta + tc$ in 1. lie on different sides of the corresponding critical arc, and hence f is sense-preserving at one pre-image and sense-reversing at the other; see Theorem 3.3 and Remark 3.5 if f is non-degenerate, and again [28, Thm. 6.7] for light harmonic mappings. □

Figure 5 (top) illustrates the effect in Theorem 4.2. The points z_1, z_2 are the pre-images of $\eta + tc$ under f, i.e., the limits of the harmonic Newton iteration for $f - (\eta + tc)$ with initial points $z_\pm$.

Remark 4.3.

1. From the proof of Lemma 2.4 we have $\operatorname{Im}(\overline{\tau(t_0)} c) > 0$, i.e., $\tau(t_0)$ and c form a right-handed ($\mathbb{R}$-)basis. Combining Theorem 4.2 with Proposition 4.1 allows to replace c by any direction d with $\operatorname{Im}(\overline{\tau(t_0)} d) > 0$ without changing the number of pre-images in $D_\varepsilon(z_0)$. More generally, if $\widetilde{\eta}$ is in the same caustic tile as $\eta + tc$ (the tile containing the tangent) and close enough to η, then $\widetilde{\eta}$ has 2 pre-images under f in $D_\varepsilon(z_0)$, and similarly in the other cases.

2. For a fold η with several pre-images in $\mathcal{C}$, the effect of Theorem 4.2 happens at all points in $f^{-1}(\{\eta\}) \cap \mathcal{C}$ simultaneously; see Example 4.4.

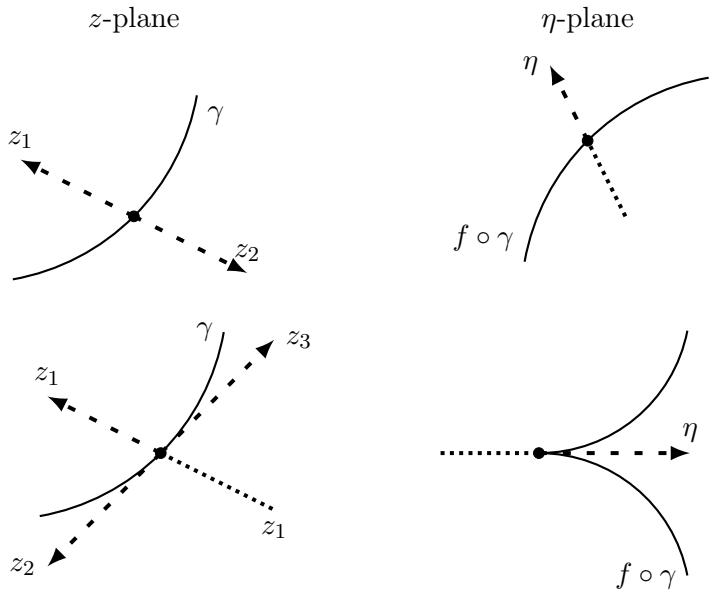


Figure 5: Behavior at a fold (top) and cusp (bottom); cf. [21, Figs. 4, 7].

3. Theorem 4.2 only covers pre-images in $D_\varepsilon(z_0)$. All other non-singular pre-images of η under f persist by Proposition 4.1, when going from η to $\eta \pm tc$, provided that $t > 0$ is sufficiently small.

When η is a cusp as in (4.1), we have a similar result, which is also based on the harmonic Newton iteration; see [33, Thm. 5.2, 2.]. For $\widetilde{\eta}$ close enough to η on one side of the caustic, there are 3 pre-images by [27, Thm 5.1], and on the other side there is only 1 pre-image by Proposition 4.1 and Theorem 4.2; see Figure 5 (bottom).

The next example illustrates the local behavior near critical points corresponding to a fold, a cusp, and a double fold, and near a point in $\mathcal{M}$.

Example 4.4. We consider the harmonic mapping $f(z) = \frac{1}{3}z^3 + \frac{1}{2}\overline{z}^2$, which is similar to the one in [28, Ex. 5.17]. Since $J_f(z) = |z|^2 - |z|^4$, we have $\mathcal{C} = \partial\mathbb{D} \cup \{0\}$ and $\mathcal{M} = \{0\}$. The caustics of f are shown in Figure 6, together with certain points $\eta_1, \ldots, \eta_6$. While "moving" η from $\eta_1 = -0.4$ to $\eta_6 = 0.9$ we reach a double fold, a point in $f(\mathcal{M})$, a fold and a cusp. The respective pre-images of η_j under f are shown in Figure 7, and have been computed with the harmonic Newton method [33]. The background is colored according to the *phase* $f_{\eta_j}/|f_{\eta_j}|$ of the shifted function $f_{\eta_j} = f - \eta_j$; see [37] for an extensive discussion of phase plots. The Poincaré index of f at z_0 corresponds to the color change on a small circle around z_0 in the positive direction. In particular, we have $\operatorname{ind}(f_{\eta_j}; z_0) = +1$ for zeros z_0 in $\mathbb{C} \setminus \overline{\mathbb{D}}$, and $\operatorname{ind}(f_{\eta_j}; z_0) = -1$ for zeros z_0 in $\mathbb{D} \setminus \{0\}$; see also Proposition 2.7. A feature is the zero $0 \in \mathcal{M}$ of f, for which $\operatorname{ind}(f; 0) = -2$ by (2.13). This reflects the

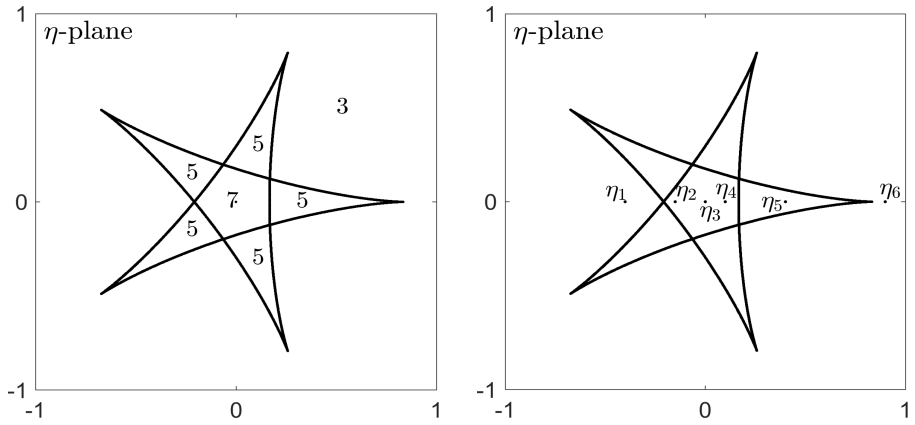


Figure 6: Caustics of $f(z) = \frac{1}{3}z^3 + \frac{1}{2}\overline{z}^2$; see Example 4.4.

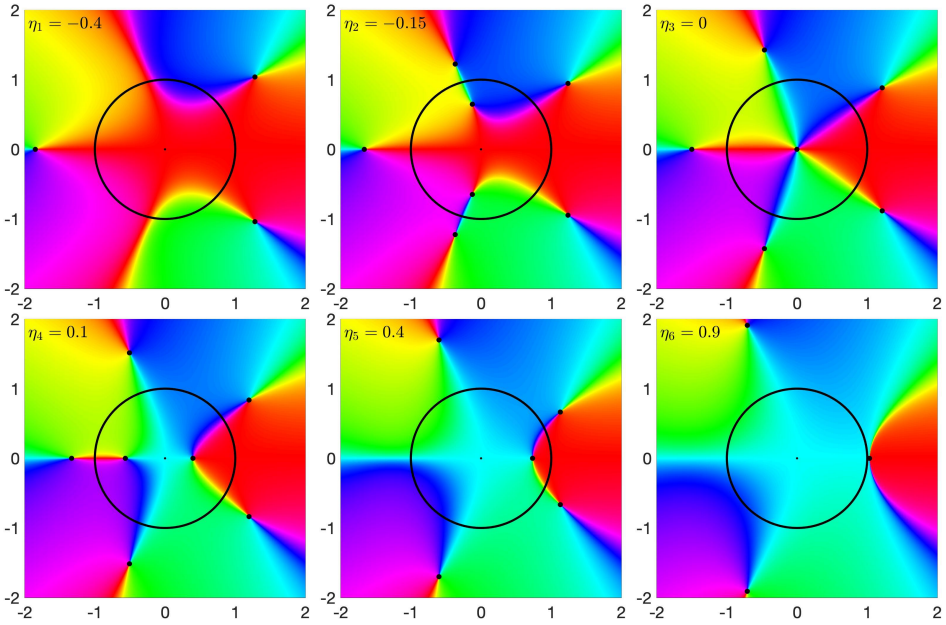


Figure 7: Phase plots of $f_{\eta_j}(z) = \frac{1}{3}z^3 + \frac{1}{2}\overline{z}^2 - \eta_j$ (see Figure 6). Black dots indicate zeros of f_{η_j}. The critical set $\mathcal{C} = \partial\mathbb{D} \cup \{0\}$ is displayed in black.

fact that two pre-images where f is sense-reversing merge together at 0; see Remark 2.8.

5 On the number of zeros of harmonic polynomials

We consider harmonic polynomials

$$f(z) = p(z) + \overline{q(z)} = \sum_{k=0}^{n} a_k z^k + \overline{\sum_{k=0}^{m} b_k z^k}, \quad n \geq m, \tag{5.1}$$

with $a_n \neq 0 \neq b_m$. These are non-degenerate if and only if $|a_n| \neq |b_n|$, where we define $b_n = 0$ for $n > m$. Such functions have at most n^2 zeros and this bound is sharp [38]. By the argument principle, f has at least n zeros, if none of them is singular. If f has fewer than n zeros, at least one has to be in $\mathcal{M}$. However, counting the zeros with their Poincaré indices as multiplicities gives again at least n zeros in total.

For $n > m \geq 1$, we study the *maximum valence* of harmonic polynomials

$$V_{n,m} = \max\{N(p(z) + \overline{q(z)}) : \deg(p) = n, \deg(q) = m\},$$

where $N(f)$ denotes the number of zeros of f. We have $V_{n,m} \leq n^2$ from [38], but the quantity $V_{n,m}$ is only known in special cases, namely $V_{n,1} = 3n - 2$ from [19, 13] and $V_{n,n-1} = n^2$ from [38]. We show in this section, that for given $n > m \geq 1$ and every $k \in \{n, n+1, \ldots, V_{n,m}\}$, there exists a harmonic polynomial (5.1) with k zeros, i.e., every number of zeros between the lower and upper bound occurs. This generalizes [7, Thm. 1.1]. More precisely, we can achieve all these numbers by just changing a_0, which is equivalent to considering the pre-images of a certain η instead of the zeros.

If η crosses a single caustic arc at a fold, the number of pre-images changes by ± 1 (η on the caustic) and ± 2 (η on the "other side" of the caustic) by Theorems 3.7 and 4.2. The key difficulty now is to handle multiple caustic arcs, i.e., caustic arcs which are the image of several different critical arcs.

Example 5.1. Consider $f(z) = \frac{1}{2}p(z)^2 + \overline{p(z)}$ with $p(z) = z^2 - 1$. Then $J_f(z) = |p'(z)|^2(|p(z)|^2 - 1)$, and $\mathcal{C} = \{z \in \mathbb{C} : |p(z)| = 1\}$ consists of the two curves $\gamma_\pm(t) = \pm\sqrt{1 + e^{-it}}$, $-\pi \leq t \leq \pi$. Since $p(\gamma_+(t)) = p(\gamma_-(t))$, the harmonic mapping f maps $\gamma_\pm$ onto the same caustic.

More generally, let γ be a closed curve with $|g'(w)/h'(w)| = 1$ on $\operatorname{trace}(\gamma)$, and let $w = p(z)$ such that $\operatorname{trace}(\gamma)$ has $k \geq 2$ disjoint pre-images under p. Then these pre-images are in the critical set of $f(z) = h(p(z)) + \overline{g(p(z))}$ and are mapped to the same caustic. In particular, $h(z) = \frac{1}{n}z^n$, $g(z) = \frac{1}{m}z^m$ with $n > m \geq 1$ provides an example of a non-degenerate harmonic polynomial with k critical curves that are mapped onto the same caustic.

Multiple caustic arcs can be eliminated by a polynomial perturbation of f. We write $\mathcal{C}_f$ and $\mathcal{C}_F$ for the critical sets of f and F, respectively.

Lemma 5.2. *Let f be a harmonic mapping, and $z_1, z_2 \in \mathcal{C}_f$, $z_1 \neq z_2$, with $f(z_1) = f(z_2)$. Then there exists a polynomial p with $\deg(p) = 3$, such that $z_1, z_2 \in \mathcal{C}_F$ for $F = f + p$, but $F(z_1) \neq F(z_2)$.*

Proof. Let $\varepsilon > 0$, and let p be the (unique) Hermite interpolation polynomial of degree 3 with $p(z_1) = \varepsilon$, $p(z_2) = -\varepsilon$, and $p'(z_1) = 0 = p'(z_2)$. We then have $J_F(z_1) = 0 = J_f(z_1)$, and the same for z_2, but $F(z_1) \neq F(z_2)$. □

Next, we show that sufficiently small perturbations do not decrease the number of non-singular zeros.

Lemma 5.3. *Let f and g be harmonic mappings, such that f has only finitely many zeros, which are all non-singular, and such that g has no singularities at the zeros of f. Then $N(f) \leq N(f + \varepsilon g)$ for all sufficiently small $\varepsilon > 0$.*

Proof. Let $z_1, \ldots, z_n$ be the zeros of f. Since non-singular zeros are isolated [12, p. 413], there exists $\delta > 0$, such that $D_\delta(z_j) \cap \mathcal{C} = \varnothing$, f and g have no other exceptional points than z_j in $\overline{D_\delta(z_j)}$ for $j = 1, \ldots, n$, and $D_\delta(z_j) \cap D_\delta(z_k) = \varnothing$ for $j \neq k$.

Define $\Gamma = \cup_{k=1}^{n} \partial D_\delta(z_k)$ and let $\varepsilon > 0$ such that

$$\varepsilon \cdot \max\{|g(z)| : z \in \Gamma\} < \min\{|f(z)| : z \in \Gamma\}.$$

Then we have for $z \in \Gamma$

$$|f(z) - (f(z) + \varepsilon g(z))| = \varepsilon |g(z)| < |f(z)|.$$

By Rouché's theorem (e.g. [32, Thm. 2.3]) and the argument principle applied on each $\partial D_\delta(z_k)$, we get

$$N(f) = \sum_{k=1}^{n} N(f; D_\delta(z_k)) \leq \sum_{k=1}^{n} N(f + \varepsilon g; D_\delta(z_k)) \leq N(f + \varepsilon g),$$

which settles the proof. □

With the Lemmas 5.2 and 5.3 we get the following result on the possible number of zeros of harmonic polynomials.

Theorem 5.4. *Let $n > m \geq 1$ and $k \in \{n, n+1, \ldots, V_{n,m}\}$. Then there exists a harmonic polynomial $f(z) = p(z) + \overline{q(z)}$ with $\deg(p) = n$ and $\deg(q) = m$, and with k zeros. Moreover, if k and n have different parity ($n - k$ is odd), then f is singular, i.e., 0 is a caustic point of f. If k and n have the same parity, then there exists a non-singular f as above.*

Proof. Let $f(z) = p(z) + \overline{q(z)}$ be a harmonic polynomial with $\deg(p) = n$, $\deg(q) = m$, and with $V_{n,m}$ zeros, which exists by the definition of $V_{n,m}$. Without loss of generality, we can assume that f has no multiple caustic arcs. Indeed, when $n = 2$ the only critical curve of f is the image of the unit circle under a Möbius transformation, and hence there are no multiple caustic arcs. If $n \geq 3$ and if f has multiple caustic arcs we resolve them by Lemma 5.2 with a polynomial perturbation of degree 3, such that no other multiple caustic arcs occur. For sufficiently small $\varepsilon > 0$, the resulting harmonic polynomial has at most $V_{n,m}$ zeros, and at least $V_{n,m}$ zeros by Lemma 5.3. This gives a harmonic polynomial with $V_{n,m}$ zeros and without multiple caustic arcs.

By Theorem 3.6, there exists an $\eta_n \in \mathbb{C}$ with $N_{\eta_n}(f) = n$. Let ϕ be a curve from η_n to 0, which intersects the caustics only in folds corresponding

to a single caustic arc. Such a curve exists since (possible) multiple caustic arcs are already resolved, and since the zeros of ψ are isolated by Lemma 2.1. Note that f is light since any $f - \eta$ has at most n^2 zeros. Then by Theorems 3.7 and 4.2, all $k = n, n+1, \ldots, V_{n,m}$ appear as number of pre-images under f for an appropriate $\eta_k \in \operatorname{trace}(\phi)$, i.e., $N_{\eta_k}(f) = k$, and hence $f - \eta_k$ is a harmonic polynomial with k zeros.

The second part follows from Theorem 3.7 and the fact that η_n can be chosen in $\mathbb{C} \setminus f(\mathcal{C})$; see Theorem 3.6. □

Remark 5.5. Let $n > m \geq 1$. By the proof of Theorem 5.4, there exists a harmonic polynomial $f(z) = p(z) + \overline{q(z)}$ with $\deg(p) = n$, $\deg(q) = m$, and $\eta_n, \ldots, \eta_{V_{n,m}} \in \mathbb{C}$, such that $f - \eta_k$ has k zeros. Moreover, $\eta_{n+1}, \eta_{n+3}, \ldots$ are on the caustics of f, and $\eta_n, \eta_{n+2}, \ldots$ can be chosen in caustic tiles.

Since $V_{n,n-1} = n^2$, we have the following corollary.

Corollary 5.6. *Let $n \geq 2$. For each $k \in \{n, n+1, \ldots, n^2\}$, there exists a harmonic polynomial as in* (5.1) *with k zeros.*

6 Outlook

A further study of the geometry of the caustics should be of interest, e.g., the number of cusps. This is an important open problem posed by Petters [29, p. 1399] for certain harmonic mappings from gravitational lensing.

While we considered harmonic mappings on the Riemann sphere (minus possible poles) in this work, also harmonic mappings in bounded domains (similar to [28]) and on more general Riemann surfaces might be of interest. We expect similar results for these domains of definition.

The results in Section 5 could probably be generalized to a broader class of harmonic mappings, e.g., non-degenerate rational harmonic mappings $f(z) = r(z) + \overline{s(z)}$, using the same approach as above. However, one would have to handle multiple caustic arcs in a different way.

Acknowledgments. We thank Jörg Liesen for several helpful comments on the manuscript. Moreover, we are grateful to the anonymous referees for many valuable comments, which lead to improvements of this work.

References

[1] L. V. Ahlfors, *Lectures on quasiconformal mappings*, D. Van Nostrand Co., Inc., Toronto, Ont.-New York-London, 1966.

[2] J. H. An and N. W. Evans, *The Chang–Refsdal lens revisited*, Monthly Notices Roy. Astronom. Soc., 369 (2006), pp. 317–334.

[3] M. B. Balk, *Polyanalytic Functions*, vol. 63 of Mathematical Research, Akademie-Verlag, Berlin, 1991.

[4] A. F. Beardon, *Complex analysis. The argument principle in analysis and topology*, John Wiley & Sons, Ltd., Chichester, 1979.

[5] C. Bénéteau and N. Hudson, *A survey on the maximal number of solutions of equations related to gravitational lensing*, in Complex analysis and dynamical systems, Trends Math., Birkhäuser/Springer, Cham, 2018, pp. 23–38.

[6] W. Bergweiler and A. Eremenko, *On the number of solutions of a transcendental equation arising in the theory of gravitational lensing*, Comput. Methods Funct. Theory, 10 (2010), pp. 303–324.

[7] P. M. Bleher, Y. Homma, L. L. Ji, and R. K. W. Roeder, *Counting zeros of harmonic rational functions and its application to gravitational lensing*, Int. Math. Res. Not. IMRN, (2014), pp. 2245–2264.

[8] B. Bollobás, *Modern graph theory*, vol. 184 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1998.

[9] D. Bshouty and A. Lyzzaik, *Problems and conjectures in planar harmonic mappings*, J. Anal., 18 (2010), pp. 69–81.

[10] J. Clunie and T. Sheil-Small, *Harmonic univalent functions*, Ann. Acad. Sci. Fenn. Ser. A I Math., 9 (1984), pp. 3–25.

[11] P. Duren, *Harmonic mappings in the plane*, vol. 156 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2004.

[12] P. Duren, W. Hengartner, and R. S. Laugesen, *The argument principle for harmonic functions*, Amer. Math. Monthly, 103 (1996), pp. 411–415.

[13] L. Geyer, *Sharp bounds for the valence of certain harmonic polynomials*, Proc. Amer. Math. Soc., 136 (2008), pp. 549–555.

[14] W. Hengartner and G. Schober, *Univalent harmonic functions*, Trans. Amer. Math. Soc., 299 (1987), pp. 1–31.

[15] P. Henrici, *Applied and computational complex analysis. Vol. 3*, Pure and Applied Mathematics (New York), John Wiley & Sons, Inc., New York, 1986.

[16] D. Khavinson, S.-Y. Lee, and A. Saez, *Zeros of harmonic polynomials, critical lemniscates, and caustics*, Complex Anal. Synerg., 4, Art. No. 2 (2018), pp. 1 – 20.

[17] D. Khavinson and G. Neumann, *On the number of zeros of certain rational harmonic functions*, Proc. Amer. Math. Soc., 134 (2006), pp. 1077–1085.

[18] D. Khavinson and G. Neumann, *From the fundamental theorem of algebra to astrophysics: a "harmonious" path*, Notices Amer. Math. Soc., 55 (2008), pp. 666–675.

[19] D. Khavinson and G. Świątek, *On the number of zeros of certain harmonic polynomials*, Proc. Amer. Math. Soc., 131 (2003), pp. 409–414.

[20] S.-Y. Lee, A. Lerario, and E. Lundberg, *Remarks on Wilmshurst's theorem*, Indiana Univ. Math. J., 64 (2015), pp. 1153–1167.

[21] J. Liesen and J. Zur, *How constant shifts affect the zeros of certain rational harmonic functions*, Comput. Methods Funct. Theory, 18 (2018), pp. 583–607.

[22] J. Liesen and J. Zur, *The maximum number of zeros of $r(z) - \overline{z}$ revisited*, Comput. Methods Funct. Theory, 18 (2018), pp. 463–472.

[23] N. G. Lloyd, *Degree theory*, Cambridge University Press, Cambridge - New York - Melbourne, 1978. Cambridge Tracts in Mathematics, No. 73.

[24] R. Luce and O. Sète, *The index of singular zeros of harmonic mappings of anti-analytic degree one*, Complex Var. Elliptic Equ., (2019). Online first publication.

[25] R. Luce, O. Sète, and J. Liesen, *Sharp parameter bounds for certain maximal point lenses*, Gen. Relativity Gravitation, 46, Art. No. 1736 (2014), pp. 1–16.

[26] R. Luce, O. Sète, and J. Liesen, *A note on the maximum number of zeros of $r(z) - \overline{z}$*, Comput. Methods Funct. Theory, 15 (2015), pp. 439–448.

[27] A. Lyzzaik, *Local properties of light harmonic mappings*, Canad. J. Math., 44 (1992), pp. 135–153.

[28] G. Neumann, *Valence of complex-valued planar harmonic functions*, Trans. Amer. Math. Soc., 357 (2005), pp. 3133–3167.

[29] A. O. Petters, *Gravity's action on light*, Notices Amer. Math. Soc., 57 (2010), pp. 1392–1409.

[30] A. O. Petters, H. Levine, and J. Wambsganss, *Singularity Theory and Gravitational Lensing*, vol. 21 of Progress in Mathematical Physics, Birkhäuser Boston, Inc., Boston, MA, 2001.

[31] J. Roe, *Winding around. The winding number in topology, geometry, and analysis*, vol. 76 of Student Mathematical Library, American Mathematical Society, Providence, RI; Mathematics Advanced Study Semesters, University Park, PA, 2015.

[32] O. Sète, R. Luce, and J. Liesen, *Perturbing rational harmonic functions by poles*, Comput. Methods Funct. Theory, 15 (2015), pp. 9–35.

[33] O. Sète and J. Zur, *A Newton method for harmonic mappings in the plane*, IMA J. Numer. Anal., 40 (2020), pp. 2777–2801.

[34] T. Sheil-Small, *Complex Polynomials*, vol. 75 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2002.

[35] T. J. Suffridge and J. W. Thompson, *Local behavior of harmonic mappings*, Complex Variables Theory Appl., 41 (2000), pp. 63–80.

[36] J. L. Walsh, *The Location of Critical Points of Analytic and Harmonic Functions*, American Mathematical Society Colloquium Publications, Vol. 34, American Mathematical Society, New York, N. Y., 1950.

[37] E. Wegert, *Visual complex functions. An introduction with phase portraits.*, Birkhäuser/Springer Basel AG, Basel, 2012.

[38] A. S. Wilmshurst, *The valence of harmonic polynomials*, Proc. Amer. Math. Soc., 126 (1998), pp. 2077–2081.

The transport of images method: computing all zeros of harmonic mappings by continuation*

Olivier Sète† Jan Zur†

Abstract

We present a continuation method to compute all zeros of a harmonic mapping f in the complex plane. Our method works without any prior knowledge of the number of zeros or their approximate location. We start by computing all solution of $f(z) = \eta$ with $|\eta|$ sufficiently large and then track all solutions as η tends to 0 to finally obtain all zeros of f. Using theoretical results on harmonic mappings we analyze where and how the number of solutions of $f(z) = \eta$ changes and incorporate this into the method. We prove that our method is guaranteed to compute all zeros, as long as none of them is singular. In our numerical example the method always terminates with the correct number of zeros, is very fast compared to general purpose root finders and is highly accurate in terms of the residual. An easy-to-use MATLAB implementation is freely available online.

Keywords: harmonic mappings, continuation method, root finding, zeros, Newton's method, critical curves and caustics.

Mathematics subject classification (2020): 65H20; 31A05; 30C55.

1 Introduction

We study the zeros of harmonic mappings, i.e., $f : \Omega \to \mathbb{C}$ with $\Delta f = 0$ on an open subset Ω of $\mathbb{C}$. These functions have a local decomposition

$$f(z) = h(z) + \overline{g(z)} \tag{1.1}$$

*A final version of this preprint manuscript appeared as follows: O. Sète and J. Zur. The transport of images method: computing all zeros of harmonic mappings by continuation. *IMA J. Numer. Anal.* (2021), doi:10.1093/imanum/drab040.

†TU Berlin, Institute of Mathematics, MA 3-3, Straße des 17. Juni 136, 10623 Berlin, Germany. {sete,zur}@math.tu-berlin.de

with analytic h, g, but are themselves non-analytic in general. Several results on the number of zeros for special classes of harmonic mappings are known in the literature, e.g., for harmonic polynomials, i.e., $f(z) = p(z)+\overline{q(z)}$, where p and q are analytic polynomials [44, 16, 9, 17], for rational harmonic mappings of the form $f(z) = r(z) - \overline{z}$, where r is a rational function [14, 5, 21, 22, 35, 34, 18, 19, 37], or certain transcendental harmonic mappings [8, 4, 13, 20].

While the above publications have a theoretical focus, we are here interested in the numerical computation of the zeros of f. A *single* zero can be computed by an iterative root finder, e.g., Newton's method on $\mathbb{R}^2$ (see [36] for a convenient formulation for harmonic mappings). The computation of *all* zeros is much more challenging. Of course, one could focus on certain regions in the complex plane and run the root finder with multiple starting points. The problem with such an approach is, however, that in general the number of zeros of a harmonic mapping is not known *a priori*. In fact, even for harmonic polynomials $f(z) = p(z) - \overline{z}$ with $\deg(p) = n \geq 2$, the number of zeros can vary between n and $3n - 2$, and each of these different numbers of zeros is attained by some harmonic polynomial; see [37, Thm. 5.4].

In this paper we present a continuation method to compute all zeros of a (non-degenerate) harmonic mapping, provided that the number of zeros is finite. We are not aware of any method for this problem in the literature that is specialized to harmonic mappings. Continuation is a general scheme that has been successfully applied to the numerical solution of systems of nonlinear equations (see, e.g., [31, Ch. 7.5] or [1]), and in particular for the solution of polynomial systems (see, e.g., [40] or [26]). Moreover, it is used in a wide range of applications, including mathematics for economics [12, Ch. 5], chemical physics [25] and astrophysics [33, Sect. 10.5]. In our continuation approach, the considered function changes only by a constant, which is sometimes called *Newton homotopy*. The overall idea is as follows. First we solve $f(z) = \eta_1$ for some $\eta_1 \in \mathbb{C}$. We then construct a sequence $\eta_2, \eta_3, \ldots$ and solve $f(z) = \eta_{k+1}$ based on the solutions of $f(z) = \eta_k$. We let the sequence $(\eta_k)_k$ tend towards the origin and eventually, after finitely many steps, we end with $\eta_n = 0$, so that solving $f(z) = \eta_n$ yields (numerically) all zeros of f.

Our method, which we call the *transport of images method*, works without any prior knowledge of the number of zeros or their approximate location in the complex plane. The main novelty is that we overcome two typical difficulties encountered in continuation methods.

1. *Finding start points:* To set up the method all solutions of $f(z) = \eta_1$ for some $\eta_1 \in \mathbb{C}$ are required. In [37, Thm. 3.6] we proved that for sufficiently large $|\eta_1|$ all solutions of $f(z) = \eta_1$ are close to the poles of f, which are usually known. Then all solutions can be computed as in [36, Sect. 4].

2. *Determine where and how the number of solutions changes:* For har-

monic mappings, the number of solutions of $f(z) = \eta_k$ depends on $\eta_k \in \mathbb{C}$; see, e.g., [37, Remark 5.5]. More precisely, the number of solutions of $f(z) = \eta_k$ and $f(z) = \eta_{k+1}$ differs by 2 if η_k and η_{k+1} are separated by a single arc of the caustics of f (curves of the critical values); see [37]. We deal with this effect as follows. For every step from η_k to η_{k+1} we construct a (minimal) set of points S_{k+1}, such that all solutions of $f(z) = \eta_{k+1}$ are obtained by applying Newton's method with initial points from S_{k+1}. This is based on results of Lyzzaik [23], Neumann [30] and the authors [37].

We prove that the transport of images method is guaranteed to compute all zeros of a non-degenerate harmonic mapping without singular zeros within a finite number of continuation steps.

Our numerical examples highlight three key features of our method: (1) It terminates with the correct number of zeros. (2) It is significantly faster than methods that are not problem adapted, such as Chebfun2[1]. (3) It is highly accurate in terms of the residual. An easy-to-use MATLAB implementation[2] is freely available online.

The name 'transport of images' method is borrowed from astrophysics. In [33, Sect. 10.5], the idea is roughly described in the context of gravitational lensing. There, the parameter η models the position of a far away light source, and its lensed images are represented by the solutions of $f(z) = \eta$ with a harmonic mapping f; see the surveys [15, 32, 3]. However, the description in [33] lacks many details and contains no further analysis.

Our paper is organized as follows. Relevant results on harmonic mappings are recalled in Section 2. In Section 3 we study the transport of images method and investigate the solution curves of $f(z) = \eta(t)$ depending on t. Key aspects of our implementation are discussed in Section 4. Section 5 contains several numerical examples that illustrate our method. We close with concluding remarks and a brief outlook in Section 6.

2 Preliminaries

We briefly present relevant results for the computation of all zeros of harmonic mappings, following the lines of [37] and [23].

2.1 Decomposition of harmonic mappings

Let $\Omega \subseteq \mathbb{C}$ be open and connected. A *harmonic mapping* on Ω is a function $f : \Omega \to \mathbb{C}$ with $\Delta f = 4\partial_{\overline{z}}\partial_z f = 0$, where ∂_z and $\partial_{\overline{z}}$ denote the *Wirtinger derivatives*. If f is harmonic in the open disk $D_r(z_0) = \{z \in \mathbb{C} : |z - z_0| < r\}$,

[1] Chebfun2, version of September 30, 2020, `www.chebfun.org`.

[2] Transport of images toolbox, `https://github.com/transportofimages/`.

it has a local decomposition

$$f(z) = h(z) + \overline{g(z)} = \sum_{k=0}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=0}^{\infty} b_k (z - z_0)^k}, \quad z \in D_r(z_0), \tag{2.1}$$

with analytic functions h and g in $D_r(z_0)$, which are unique up to an additive constant; see [7, p. 412] or [6, p. 7]. If f is harmonic in the punctured disk $D = \{z \in \mathbb{C} : 0 < |z - z_0| < r\}$, i.e., z_0 is an isolated singularity of f, then

$$f(z) = \sum_{k=-\infty}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=-\infty}^{\infty} b_k (z - z_0)^k} + 2A \log|z - z_0|, \quad z \in D; \tag{2.2}$$

see [41, 10, 2]. The point z_0 is a *pole* of f if $\lim_{z \to z_0} f(z) = \infty$, and an *essential singularity* if the limit does not exist in $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$; see [38, p. 44], [41, Def. 2.1] or [37, Def. 2.9]. It is possible that f has a pole at z_0, but both series in (2.2) have an essential singularity at z_0, as the example $f(z) = z^{-2} + \cos(i/z) + \overline{\cos(i/z)}$ from [41, Ex. 2.3] shows. It is also possible that f has an essential singularity at z_0, but both series have a pole at z_0, as shown by $f(z) = z^{-1} + \overline{z}^{-1}$.

The next proposition is the analog of the well-known facts that a function holomorphic on $\widehat{\mathbb{C}}$ is constant and a function meromorphic on $\widehat{\mathbb{C}}$ is rational.

Proposition 2.1. *A harmonic mapping f on $\widehat{\mathbb{C}} \setminus \{z_1, \ldots, z_m\}$ with poles at $z_1, \ldots, z_m \in \widehat{\mathbb{C}}$, at which the principal parts of both series in the decomposition* (2.2) *have only finitely many terms, has the form*

$$f(z) = r(z) + \overline{s(z)} + \sum_{j=1}^{m} 2A_j \log|z - z_j|, \tag{2.3}$$

where $A_1, \ldots, A_m \in \mathbb{C}$, and r and s are rational functions that can have poles only at $z_1, \ldots, z_m$. If some $z_j = \infty$, then $\log|z|$ replaces $\log|z - z_j|$. In particular, if f is harmonic on $\widehat{\mathbb{C}}$, then f is constant.

Proof. First, let f be a harmonic mapping on $\widehat{\mathbb{C}}$. Then $\partial_z f$ is analytic on $\widehat{\mathbb{C}}$ and thus constant [43, Thm. 3.5.8]. Let $h' = \partial_z f$, then $h(z) = a_1 z + a_0$ is analytic. Let $g = \overline{f - h}$. Since $\partial_{\overline{z}} g = \overline{\partial_z f - h'} = 0$, g is analytic on $\widehat{\mathbb{C}}$ and hence constant, say $g(z) = b_0$, and we have $f(z) = a_1 z + a_0 + \overline{b}_0$. Since f is harmonic at ∞ we have $a_1 = 0$ and f is constant. Next, let f be harmonic with poles at $z_1, \ldots, z_m$. If $z_j \neq \infty$ the decomposition (2.2) has the form

$$f(z) = \sum_{k=-n}^{\infty} (a_k (z - z_j)^k + \overline{b_k (z - z_j)^k}) + 2A_j \log|z - z_j|,$$

and we set $f_j(z) = \sum_{k=-n}^{-1} (a_k (z - z_j)^k + \overline{b_k (z - z_j)^k}) + 2A_j \log|z - z_j|$ and similarly if $z_j = \infty$. Then $f - f_1 - \ldots - f_m$ has removable singularities at $z_1, \ldots, z_m$, and can be extended to a harmonic mapping in $\widehat{\mathbb{C}}$ (see [37, Sect. 2.2], or also [11, Thm. 15.3d]), which is constant. □

2.2 Critical set and caustics of harmonic mappings

The critical points of a harmonic mapping $f : \Omega \to \mathbb{C}$, i.e., the points where the *Jacobian* of f,

$$J_f(z) = |\partial_z f(z)|^2 - |\partial_{\overline{z}} f(z)|^2, \tag{2.4}$$

vanishes, form the *critical set* of f,

$$\mathcal{C} = \{z \in \Omega : J_f(z) = 0\}. \tag{2.5}$$

By Lewy's theorem [6, p. 20] f is locally univalent at $z \in \Omega$ if and only if $z \notin \mathcal{C}$. Here and in the following we assume that $\partial_z f$ and $\partial_{\overline{z}} f$ do not vanish identically. (Otherwise f would be analytic or anti-analytic, which is not the focus of this paper.) Then $\partial_z f$ has only isolated zeros in Ω, and the critical set consists of isolated points and a level set of the *second complex dilatation* of f,

$$\omega(z) = \frac{\overline{\partial_{\overline{z}} f(z)}}{\partial_z f(z)}, \tag{2.6}$$

which is a meromorphic function in Ω. We assume that removable singularities of ω in Ω are removed. Let

$$\mathcal{M} = \{z \in \mathcal{C} : |\omega(z)| \neq 1\} \tag{2.7}$$

$$= \{z \in \Omega : \partial_z f(z) = \partial_{\overline{z}} f(z) = 0 \text{ and } \lim_{\zeta \to z} |\omega(\zeta)| \neq 1\}, \tag{2.8}$$

see [37, Sect. 2.1], which consists of isolated points of $\mathcal{C}$ [23, Lem. 2.2]. Then

$$\mathcal{C} \setminus \mathcal{M} = \{z \in \Omega : |\omega(z)| = 1\} \tag{2.9}$$

is a level set of ω. If $J_f \not\equiv 0$, then $\mathcal{C} \setminus \mathcal{M}$ consists of analytic curves. These intersect in $z \in \mathcal{C} \setminus \mathcal{M}$ if, and only if, $\omega'(z) = 0$. At $z \in \mathcal{C} \setminus \mathcal{M}$ with $\omega'(z) \neq 0$, the equation

$$\omega(\gamma(t)) = e^{it}, \quad t \in I \subseteq \mathbb{R}, \tag{2.10}$$

implicitly defines a local analytic parametrization $z = \gamma(t)$ of $\mathcal{C} \setminus \mathcal{M}$.

Definition 2.2. We call the set of critical values of f, i.e., $f(\mathcal{C})$, the *caustics* or the *set of caustic points*. The caustics induce a partition of $\mathbb{C} \setminus f(\mathcal{C})$, and we call a connected component $A \subseteq \mathbb{C} \setminus f(\mathcal{C})$ with $\partial A \subseteq f(\mathcal{C})$ a *caustic tile*.

The next lemma characterizes a tangent vector to the caustics; see [37, Lem. 2.1] or [23, Lem. 2.3].

Lemma 2.3. *Let f be a harmonic mapping, $z_0 \in \mathcal{C} \setminus \mathcal{M}$ with $\omega'(z_0) \neq 0$ and let $z_0 = \gamma(t_0)$ with the parametrization* (2.10)*. Then $f \circ \gamma$ is a parametrization of a caustic and the corresponding tangent vector at $f(z_0)$ is*

$$\tau(t_0) = \frac{d}{dt}(f \circ \gamma)(t_0) = e^{-it_0/2}\psi(t_0) \tag{2.11}$$

with

$$\psi(t) = 2\,\mathrm{Re}(e^{it/2}\,\partial_z f(\gamma(t))\,\gamma'(t)). \tag{2.12}$$

In particular, the rate of change of the argument of the tangent vector is

$$\frac{d}{dt}\arg(\tau(t))\big|_{t=t_0} = -\frac{1}{2} \tag{2.13}$$

at points where $\psi(t_0) \neq 0$, i.e., the curvature of the caustics is constant with respect to the parametrization $f \circ \gamma$. Moreover, ψ has either only finitely many zeros, or is identically zero, in which case $f \circ \gamma$ is constant.

Definition 2.4 ([37, Def. 2.2])**.** In the notation of Lemma 2.3 we call $f(\gamma(t_0))$ a *fold caustic point* or simply a *fold*, if the tangent $\tau(t_0)$ is non-zero, i.e., $\psi(t_0) \neq 0$, and a *cusp* if ψ has a zero with a sign change at t_0.

This classification of caustic points is not complete but sufficient for our needs. Fold points form open arcs of the caustic curves, since ψ is continuous. At a cusp the argument of the tangent vector changes by $+\pi$. The cusps, the curvature of the caustics and the four caustic tiles are apparent in Figure 1 (left).

2.3 Pre-images under non-degenerate harmonic mappings

The position of $\eta \in \mathbb{C}$ with respect to the caustics affects the number of pre-images of η under f. The following wide class of harmonic mappings was introduced in [37, Def. 3.1].

Definition 2.5. We call a harmonic mapping f *non-degenerate* on $\widehat{\mathbb{C}}$ if

1. f is defined in $\widehat{\mathbb{C}}$ except at finitely many poles $z_1, \ldots, z_m \in \widehat{\mathbb{C}}$,
2. at each pole $z_j \in \mathbb{C}$ of f, the decomposition (2.2) of f has the form

$$f(z) = \sum_{k=-n}^{\infty} a_k (z - z_j)^k + \overline{\sum_{k=-n}^{\infty} b_k (z - z_j)^k} + 2A_j \log|z - z_j| \tag{2.14}$$

   with $n \geq 1$ and $|a_{-n}| \neq |b_{-n}|$, and if $z_j = \infty$ is a pole of f, then

$$f(z) = \sum_{k=-\infty}^{n} a_k z^k + \overline{\sum_{k=-\infty}^{n} b_k z^k} + 2A_j \log|z|, \quad \text{for } |z| > R, \tag{2.15}$$

   with $n \geq 1$ and $|a_n| \neq |b_n|$, and $R > 0$,
3. the critical set of f is bounded in $\mathbb{C}$.

The intensively studied (see references in the introduction) rational harmonic mappings $r(z) - \overline{z}$ and harmonic polynomials $p(z) + \overline{q(z)}$ are non-degenerate if, and only if, $\lim_{z\to\infty}|r(z)/z| \neq 1$ and $\lim_{z\to\infty}|p(z)/q(z)| \neq 1$, respectively, since ∞ is the only pole of $p(z)$, $q(z)$ and z.

A non-degenerate harmonic mapping f has a global decomposition (2.3), and

$$\partial_z f(z) = r'(z) + \sum_{j=1}^{m} \frac{A_j}{z - z_j}, \quad \overline{\partial_{\overline{z}} f(z)} = s'(z) + \sum_{j=1}^{m} \frac{\overline{A}_j}{z - z_j}, \tag{2.16}$$

are rational functions. By item 2., the poles are not accumulation points of the critical set [41, Lem. 2.2, 2.3]. Hence, $\mathcal{C}$ and $f(\mathcal{C})$ are compact sets in $\mathbb{C}$. Item 3. implies that $J_f \not\equiv 0$. If $\partial_z f \equiv 0$, then $\overline{\partial_{\overline{z}} f} \not\equiv 0$ (since $J_f \not\equiv 0$), and $\mathcal{C}$ consists of the finitely many zeros of the rational function $\overline{\partial_{\overline{z}} f}$. If $\partial_z f \not\equiv 0$, then $\omega = \overline{\partial_{\overline{z}} f}/\partial_z f$ is rational, $\mathcal{M}$ has only finitely many points, and $\mathcal{C} \setminus \mathcal{M}$ consists of the $\deg(\omega)$ many pre-images of the unit circle under ω. We can parametrize $\mathcal{C} \setminus \mathcal{M}$ with analytic closed curves according to (2.10), such that every $z \in \mathcal{C} \setminus \mathcal{M}$ with $\omega'(z) \neq 0$ belongs to exactly one curve; see [37, Seect. 3.1]. We call these curves the *critical curves* of f and denote the set of all of them by crit. The critical curves and their images under f both have finite total length since their parametrization is piecewise C^∞.

Let $P(f) = n_1 + \ldots + n_m$ be the sum of the orders of the poles $z_1, \ldots, z_m$ in (2.14) or (2.15). Then $N_\eta(f) = |\{z \in \widehat{\mathbb{C}} : f(z) = \eta\}|$, the number of pre-images of η under f, relates to $P(f)$ and the winding number $n(f \circ \gamma; \eta)$ of the caustics about η as follows.

Theorem 2.6 ([37, Thms. 3.4, 3.7])**.** *Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$. Then for a point $\eta \in \mathbb{C} \setminus f(\mathcal{C})$ we have*

$$N_\eta(f) = P(f) + 2 \sum_{\gamma \in \mathrm{crit}} n(f \circ \gamma; \eta). \tag{2.17}$$

In particular, the number of pre-images of $\eta \in \mathbb{C} \setminus f(\mathcal{C})$ under f is finite. Moreover, if $\eta_1, \eta_2 \in \mathbb{C} \setminus f(\mathcal{C})$ then

$$N_{\eta_2}(f) = N_{\eta_1}(f) + 2 \sum_{\gamma \in \mathrm{crit}} \big(n(f \circ \gamma; \eta_2) - n(f \circ \gamma; \eta_1)\big), \tag{2.18}$$

and we have:

1. *If η_1 and η_2 are in the same caustic tile then the number of pre-images under f is the same, i.e., $N_{\eta_2}(f) = N_{\eta_1}(f)$.*
2. *If η_1 and η_2 are separated by a single caustic arc $f \circ \gamma$ then the number of pre-images under f differs by 2, i.e., $N_{\eta_2}(f) = N_{\eta_1}(f) \pm 2$.*

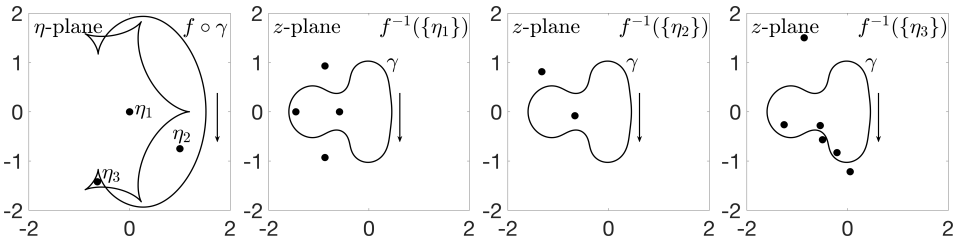


Figure 1: Harmonic mapping (2.19). η-plane: the caustic $f \circ \gamma$ of f and the points η_j. z-plane: pre-images of η_j under f. The arrows indicate the orientation of the critical curve γ and the caustic $f \circ \gamma$.

Example 2.7 ([37, Ex. 3.10])**.** We study the non-degenerate harmonic mapping

$$f(z) = z^2 + \overline{z^{-1} + (z+1)^{-1}} + 2\log|z|; \tag{2.19}$$

see Figure 1. By Lemma 2.3 the curvature of a caustic is always negative. Consequently, the winding numbers of the only caustic $f \circ \gamma$ about η_1, η_2 and η_3 are 0, -1 and 1 respectively. By Theorem 2.6 we have

$$\begin{aligned} N_{\eta_1}(f) &= P(f) + 2n(f \circ \gamma; \eta_1) = 4 + 0 = 4, \\ N_{\eta_2}(f) &= N_{\eta_1}(f) + 2(n(f \circ \gamma; \eta_2) - n(f \circ \gamma; \eta_1)) = 4 - 2 = 2, \\ N_{\eta_3}(f) &= N_{\eta_1}(f) + 2(n(f \circ \gamma; \eta_3) - n(f \circ \gamma; \eta_1)) = 4 + 2 = 6, \end{aligned}$$

as we see in Figure 1.

3 The transport of images method

We compute all zeros of a non-degenerate harmonic mapping f by continuation, provided that no zero of f is singular, or equivalently, if 0 belongs to the caustics of f. A zero z_0 is *singular*, if $J_f(z_0) = 0$ holds. Recall from the introduction that we successively compute all solutions of $f(z) = \eta_k$ for a sequence $\eta_1, \eta_2, \ldots$ that tends to zero. During this procedure the number of solutions may change, depending on the positions of the points η_k relative to the caustics of f; see Theorem 2.6. We start by computing all solutions of $f(z) = \eta_1$ for sufficiently large $|\eta_1|$ from the local decomposition (2.2) of f near the poles. Given all solutions of $f(z) = \eta_k$ we determine an $\eta_{k+1} \in \mathbb{C}$ and a set $S_{k+1} \subset \mathbb{C}$ (prediction) such that the following holds: (1) for every solution $z_* \in \mathbb{C}$ of $f(z) = \eta_{k+1}$ there exists a $z_0 \in S_{k+1}$ such that the sequence $(z_j)_{j\in\mathbb{N}}$ of Newton iterates for $f - \eta_{k+1}$ converges to z_* (correction); (2) the number of solutions of $f(z) = \eta_{k+1}$ coincides with the number of elements in S_{k+1}, i.e., S_{k+1} is minimal. When this prediction-correction scheme reaches $\eta_n = 0$ we obtain *all* zeros of f. This leads to the following algorithm.

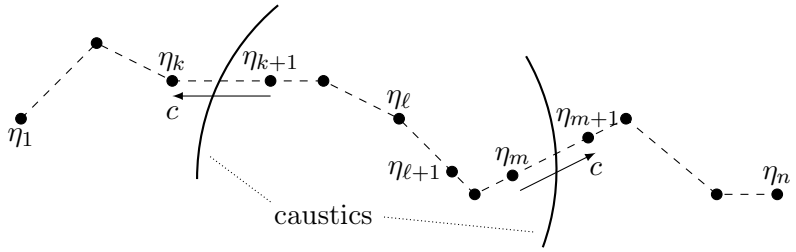


Figure 2: A path from η_1 to η_n with the three kinds of steps in the transport of images method: (1) $f(z) = \eta_{k+1}$ has two solutions fewer than $f(z) = \eta_k$, (2) the number of solutions for η_ℓ and $\eta_{\ell+1}$ coincides and (3) $f(z) = \eta_{m+1}$ has two solutions more than $f(z) = \eta_m$.

The transport of images method

1. *Initial phase:* compute all solutions of $f(z) = \eta_1$.
2. *Transport phase:* while $\eta_k \neq 0$ do
   (a) *Prediction:* choose $\eta_{k+1} \in \mathbb{C}$ and construct a minimal set of initial points S_{k+1} from the solutions of $f(z) = \eta_k$ and from the caustics.
   (b) *Correction:* apply Newton's method to $f - \eta_{k+1}$ with the set of initial points S_{k+1} to get all solutions of $f(z) = \eta_{k+1}$.

The points $\eta_1, \eta_2, \ldots, \eta_n = 0$ conceptually form a path in $\mathbb{C}$ (see Figure 2). From this perspective the transport of images method discretizes the tracing of the solution curves of $f(z) = \eta(t)$, where the right hand side is a path $\eta : [a, b] \to \mathbb{C}$. We describe the homotopy curves for a general harmonic mapping in Section 3.4 in full detail.

3.1 Newton's method as corrector

For a harmonic mapping $f : \Omega \to \mathbb{C}$ the *harmonic Newton map* is

$$H_f : \Omega \setminus \mathcal{C} \to \mathbb{C}, \quad H_f(z) = z - \frac{\overline{h'(z)} f(z) - \overline{g'(z) f(z)}}{|h'(z)|^2 - |g'(z)|^2}, \tag{3.1}$$

where $f = h + \overline{g}$ is a local decomposition (2.1) and the *harmonic Newton iteration* with initial point $z_0 \in \Omega \setminus \mathcal{C}$ is

$$z_k = H_f(z_{k-1}) = H_f^k(z_0), \quad k \geq 1. \tag{3.2}$$

We write, as usual, $H_f^k(z) = H_f(H_f^{k-1}(z))$ for $k \geq 1$ and $H_f^0(z) = z$. The iteration (3.2) is a complex formulation of the classical Newton iteration in $\mathbb{R}^2$ [36, Sect. 3.1] and hence inherits all its properties, e.g., local quadratic convergence.

To compute solutions of $f(z) = \eta$ with $\eta \in \mathbb{C}$ we apply the iteration (3.2) to $f - \eta$. The limit of the harmonic Newton iteration with initial point z_0,

$$H^{\infty}_{f-\eta}(z_0) = \lim_{k\to\infty} H^{k}_{f-\eta}(z_0), \tag{3.3}$$

exists for all z_0 in a neighborhood of a solution $z_* \notin \mathcal{C}$ of $f(z) = \eta$ by the local convergence property of Newton's method. If $z_* = H^{\infty}_{f-\eta}(z_0)$ holds we say that z_* *attracts* z_0 under $H_{f-\eta}$ or that z_0 *is attracted* by z_* under $H_{f-\eta}$.

Definition 3.1. For a harmonic mapping f we call a set S a *(minimal) prediction set* of $f^{-1}(\{\eta\})$ if every solution of $f(z) = \eta$ attracts exactly one point in S under $H_{f-\eta}$, i.e., if $H^{\infty}_{f-\eta}(S) = f^{-1}(\{\eta\})$, and $|S| = |f^{-1}(\{\eta\})|$, where $|X|$ denotes the cardinality of a set X.

For a prediction set S the map $H^{\infty}_{f-\eta}$ is a bijection from S to $f^{-1}(\{\eta\})$. In this sense every point $z_0 \in S$ represents exactly one solution of $f(z) = \eta$. In [39] the point z_0 is called an *approximate zero* of $f - \eta$ if the Newton iterates additionally satisfy $|z_k - z_{k-1}| \le (\frac{1}{2})^{2^{k-1}-1}|z_1 - z_0|$.

3.2 Initial phase

For non-degenerate harmonic mappings the initial phase can be realized as follows. For sufficiently large $|\eta|$ all solutions of $f(z) = \eta$ are close to the poles of f and, in the notation (2.14) or (2.15), there are exactly n solutions close to the pole z_j [37, Thm. 3.6]. We compute these solutions with the harmonic Newton iteration and the initial points from the next theorem. A similar result holds if ∞ is a pole of f, generalizing [36, Cor. 4.5]. All initial points together give a prediction set of $f^{-1}(\{\eta\})$.

Theorem 3.2. *Let f be harmonic in $D = \{z \in \mathbb{C} : 0 < |z - z_0| < r\}$ with*

$$f(z) = \sum_{k=-n}^{\infty} a_k(z - z_0)^k + \overline{\sum_{k=-n}^{\infty} b_k(z - z_0)^k} + 2A\log|z - z_0|, \quad z \in D,$$

where $n \ge 1$ and $|a_{-n}| \ne |b_{-n}|$. Suppose that $c = \eta - (a_0 + \overline{b}_0) \ne 0$ and let $\zeta_1, \ldots, \zeta_n$ be the n solutions of

$$(\zeta - z_0)^n = \frac{|a_{-n}|^2 - |b_{-n}|^2}{\overline{a}_{-n}c - \overline{b}_{-n}\overline{c}}. \tag{3.4}$$

We then have for sufficiently large $|c|$, i.e., for sufficiently large $|\eta|$:

1. *There exist exactly n solutions of $f(z) = \eta$ near z_0.*
2. *The set $S = \{\zeta_1, \ldots, \zeta_n\}$ is a prediction set for the solutions in 1.*

Proof. If $A = 0$ then $f - \eta$ has n distinct zeros near z_0, which satisfy 2. by [36, Thm. 4.3]. For $A \ne 0$ the proof closely follows the proof of [36, Thm. 4.3] if one replaces h' by $\partial_z f$, and g' by $\overline{\partial_{\bar z} f}$. As in the proof of [37, Thm. 3.6], Rouché's theorem (e.g. [35, Thm. 2.3]) implies that $f - \eta$ cannot have more than n zeros in a neighborhood of z_0. □

3.3 Transport phase

We analyze the prediction-correction scheme of the transport phase and prove that the transport phase can be realized with a finite number of steps. We begin with a single step and investigate how all solutions of $f(z) = \eta_2$ can be determined, given all solutions of $f(z) = \eta_1$. First we prove that a solution of $f(z) = \eta_1$ that is not in $\mathcal{C}$ is attracted by a solution of $f(z) = \eta_2$ under the harmonic Newton map $H_{f-\eta_2}$ if $|\eta_2 - \eta_1|$ is small enough.

Lemma 3.3. *Let $f : \Omega \to \mathbb{C}$ be a harmonic mapping, $z_0 \in \Omega \setminus \mathcal{C}$ and $\eta_1 = f(z_0)$. Then there exist $\varepsilon > 0$ and $\delta > 0$ such that for each $\eta_2 \in D_\varepsilon(\eta_1)$ there exists a unique $z_* \in D_\delta(z_0)$ with $f(z_*) = \eta_2$. Moreover, z_* attracts z_0 under the harmonic Newton map $H_{f-\eta_2}$, i.e., $H^\infty_{f-\eta_2}(z_0) = z_*$.*

Proof. Let $\zeta \in \Omega \setminus \mathcal{C}$ and consider $f - f(\zeta)$. The Newton–Kantorovich theorem yields a radius $\rho(\zeta) > 0$ so that ζ attracts all points in $D_{\rho(\zeta)}(\zeta)$ under the harmonic Newton map of $f - f(\zeta)$; see [36, Thm. 2.2, Sect. 4] and references therein. (The basin of attraction of ζ might be larger.) Note that ρ depends continuously on ζ.

There exist open neighborhoods $U \subseteq \Omega \setminus \mathcal{C}$ of z_0 and V of η_1 such that $f : U \to V$ is a diffeomorphism (inverse function theorem). Let $r > 0$ with $K = \overline{D_r(z_0)} \subseteq U$ and let $m = \min_{z \in K} \rho(z)$. Then $m > 0$ since K is compact and ρ is continuous with $\rho(z) > 0$ for $z \in U$. Let $0 < \delta \leq \min\{r, m\}$. Then $z_* \in D_\delta(z_0)$ satisfies $|z_* - z_0| < \delta \leq m \leq \rho(z_*)$, hence $H^\infty_{f-f(z_*)}(z_0) = z_*$. Finally, there exists $\varepsilon > 0$ with $D_\varepsilon(\eta_1) \subseteq f(D_\delta(z_0))$. □

By Theorem 2.6, two points in the same caustic tile have the same number of pre-images. Moreover, $\eta \in \mathbb{C} \setminus f(\mathcal{C})$ has only finitely many pre-images, and these are in $\widehat{\mathbb{C}} \setminus \mathcal{C}$. Thus, we can apply Lemma 3.3 to all solutions simultaneously to obtain the next theorem.

Theorem 3.4. *Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$ and let $\eta_1 \in \mathbb{C} \setminus f(\mathcal{C})$. Then there exists an $\varepsilon > 0$ such that $S = f^{-1}(\{\eta_1\})$ is a prediction set of $f^{-1}(\{\eta_2\})$ for all $\eta_2 \in D_\varepsilon(\eta_1) \subseteq \mathbb{C} \setminus f(\mathcal{C})$.*

By Theorem 2.6, the number of pre-images of η_1, η_2 differs by 2 if η_1, η_2 are in adjacent caustic tiles separated by a single caustic arc. The two additional solutions can be computed with the next theorem, provided that the step from η_1 to η_2 crosses the caustics in a specific direction.

Theorem 3.5 ([37, Thm. 4.2])**.** *Let $f : \Omega \to \mathbb{C}$ be a harmonic mapping and $z_0 \in \mathcal{C} \setminus \mathcal{M}$, such that $\eta = f(z_0)$ is a fold caustic point. Moreover, let*

$$f(z) = \sum_{k=0}^{\infty} a_k (z - z_0)^k + \overline{\sum_{k=0}^{\infty} b_k (z - z_0)^k} \quad \text{and } c = -\left(\frac{a_2 \overline{b}_1}{a_1} + \frac{\overline{b}_2 a_1}{\overline{b}_1} \right). \tag{3.5}$$

Then, for each sufficiently small $\varepsilon > 0$, there exists a $\delta > 0$ such that for all $0 < t < \delta$:

1. *$f(z) = \eta - tc$ has no solution in $D_\varepsilon(z_0)$,*

2. *$f(z) = \eta$ has exactly one solution in $D_\varepsilon(z_0)$,*

3. *$f(z) = \eta + tc$ has exactly two solutions in $D_\varepsilon(z_0)$.*

Moreover, each solution in 3. attracts one of the points

$$z_\pm = z_0 \pm i\sqrt{t\bar{b}_1/a_1} \tag{3.6}$$

under the harmonic Newton map $H_{f-(\eta+tc)}$ if $t > 0$ is sufficiently small.

Note that f is *light* (i.e., $f^{-1}(\{\eta\})$ is either totally disconnected or empty for all $\eta \in \mathbb{C}$) in a neighborhood of z_0 if $f(z_0)$ is a fold [37, Rem. 2.3]; hence we omit this assumption from [36, Thm. 4.2]. We emphasize that c is not necessarily a normal vector to the caustics.

With Theorem 3.5, we construct a prediction set of $f^{-1}(\{\eta_2\})$ if the step from η_1 to η_2 crosses a caustic at a *simple fold* η, i.e., $|f^{-1}(\{\eta\}) \cap \mathcal{C}| = 1$.

Theorem 3.6. *Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$. Let $\eta_1, \eta_2 \in \mathbb{C} \setminus f(\mathcal{C})$ such that there exists exactly one simple fold η and no other caustic point on the line segment from η_1 to η_2, and such that $\arg(\eta_2 - \eta_1) = \arg(\pm c)$, where c is defined as in Theorem 3.5. Then for small enough $|\eta_2 - \eta_1|$,*

$$S = \begin{cases} f^{-1}(\{\eta_1\}) \cup \{z_\pm\}, & \text{if } \arg(\eta_2 - \eta_1) = \arg(+c), \\ f^{-1}(\{\eta_1\}) \setminus H^\infty_{f-\eta_1}(\{z_\pm\}), & \text{if } \arg(\eta_2 - \eta_1) = \arg(-c), \end{cases} \tag{3.7}$$

is a prediction set of $f^{-1}(\{\eta_2\})$, with $z_\pm$ from (3.6).

Proof. Let $\eta = f(z_0)$ be the unique simple fold on the line segment between η_1 and η_2. The step from η_1 to η_2 produces either two additional or two fewer solutions in a neighborhood of z_0, depending on c; see Theorem 3.5. The global number of solutions changes accordingly; see Theorem 2.6.

In the case of two additional solutions, i.e., if $\arg(\eta_2 - \eta_1) = \arg(+c)$, the set S contains the solutions of $f(z) = \eta_1$, and the points $z_\pm$, hence $|S| = |f^{-1}(\{\eta_2\})|$ for sufficiently small $|\eta_2 - \eta_1|$. Let $t_1, t_2 > 0$ such that $\eta_1 = \eta - t_1 c$ and $\eta_2 = \eta + t_2 c$. In particular, we have $t_j \leq |(\eta_2 - \eta_1)/c|$. In the disk $D_\varepsilon(z_0)$ from Theorem 3.5, $f(z) = \eta_1$ has no solution and $f(z) = \eta_2$ has exactly two solutions, which attract z_+ and z_-, respectively, if $|\eta_2 - \eta_1|$ and hence t_2 are small enough. Since η_1 is not a caustic point, all solutions of $f(z) = \eta_1$ are not in $\mathcal{C}$. Hence, every solution of $f(z) = \eta_2$ outside $D_\varepsilon(z_0)$ attracts exactly one solution of $f(z) = \eta_1$ by Lemma 3.3, if $|\eta_2 - \eta_1|$ is sufficiently small. Together, S is a prediction set of $f^{-1}(\{\eta_2\})$.

In the case of two fewer solutions, i.e., if $\arg(\eta_2 - \eta_1) = \arg(-c)$, we remove two points from $f^{-1}(\{\eta_1\})$ to obtain a prediction set. To determine

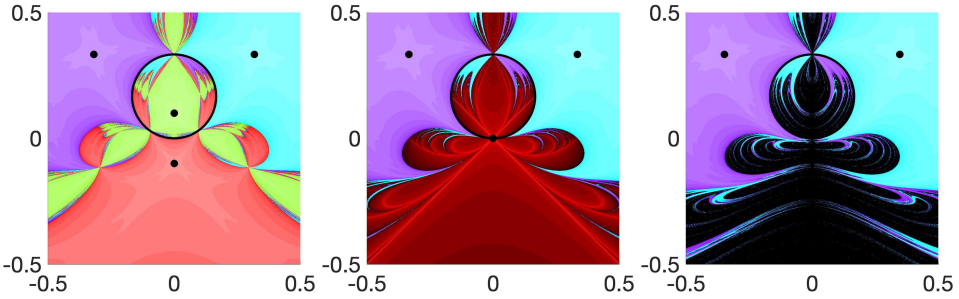


Figure 3: Solutions (black dots) of $f(z) = z + 2iz + \overline{z + iz^2} = tc$ with basins of attraction for $z_0 = 0$, $c = -i$ and $t = 0.01$, $t = 0$, $t = -0.01$ (from left to right); see Remark 3.8. The black circle is the critical set of f.

these points we consider the reversed step from η_2 to η_1. This step gives the two additional solutions $H^{\infty}_{f-\eta_1}(z_\pm)$, which we remove from $f^{-1}(\{\eta_1\})$ to get S. Each solution of $f(z) = \eta_2$ attracts exactly one point in S, as above. □

Remark 3.7. Let $\eta \in f(\mathcal{C})$ be a (multiple) fold caustic point, i.e., for each $z_j \in \{z_1, \dots, z_m\} = f^{-1}(\{\eta\}) \cap \mathcal{C}$ the tangent τ_j from (2.11) is non-zero. Then, for every $d \in \mathbb{C}$ with $\mathrm{Im}(\overline{\tau}_j d) \neq 0$ for $j = 1, \dots, m$, the effect of Theorem 3.5 happens at all points $z_1, \dots, z_m$ simultaneously. By [37, Rem. 4.3], $f(z) = \eta + td$ has 2 solutions in $D_\varepsilon(z_j)$ and $f(z) = \eta - td$ has no solutions in $D_\varepsilon(z_j)$ if $\mathrm{Im}(\overline{\tau}_j d) > 0$ and $t > 0$ is sufficiently small. If $\mathrm{Im}(\overline{\tau}_j d) < 0$ then $f(z) = \eta \pm td$ swap their roles. However, it is not guaranteed that the additional solutions attract the points $z_\pm(z_j)$ if $\arg(d) \neq \arg(\pm c(z_j))$, with $z_\pm$ and c from Theorem 3.5.

Remark 3.8. The basins of attraction of the two solutions of $f(z) = \eta + tc$ in $D_\varepsilon(z_0)$ ($t > 0$) merge ($t = 0$) and disappear ($t < 0$) with the solutions. This is illustrated in Figure 3. Points that are attracted by the same solution have the same color. The darker the shading, the more iterations are needed until (numerical) convergence. Points where the iteration does not converge (numerically) are colored in black. This highlights why it is important to remove the two points $H^{\infty}_{f-\eta_1}(z_\pm)$ from $f^{-1}(\{\eta_1\})$ in (3.7) for practical computations.

We illustrate Theorems 3.4 and 3.6 in Figure 4. First, consider the step from η_1 to η_2 and the dynamics of $H_{f-\eta_2}$ (middle plot). Every basin of attraction contains exactly one element of $S = f^{-1}(\{\eta_1\})$, i.e., S is a prediction set of $f^{-1}(\{\eta_2\})$. Some elements in S are close to the boundary of the respective basins, such that S may not be a prediction set of $f^{-1}(\{\eta_2\})$ if η_2 is slightly perturbed. Since the step from η_2 to η_3 crosses a caustic, the number of solutions of $f(z) = \eta_2$ and $f(z) = \eta_3$ differ. According to Theorem 3.6 $S = f^{-1}(\{\eta_2\}) \cup \{z_\pm\}$ with the points $z_\pm$ from (3.6) is a prediction set of $f^{-1}(\{\eta_3\})$, shown in the right plot.

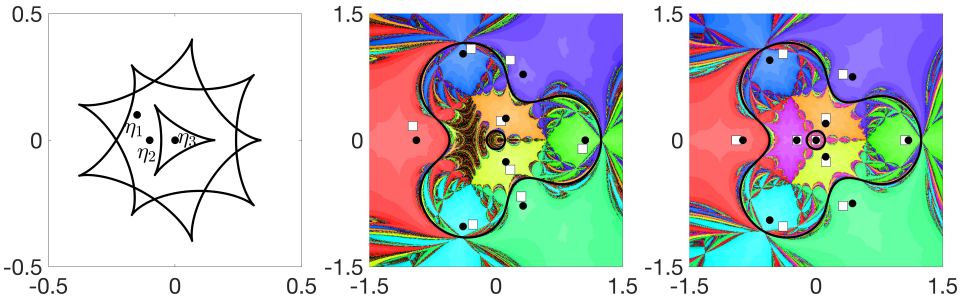


Figure 4: Illustration of Theorems 3.4 and 3.6 for $f(z) = z - \overline{z^2/(z^3 - 0.6^3)}$ (see (5.2) below). Left: caustics and η_1, η_2, η_3. Middle: prediction set $S = f^{-1}(\{\eta_1\})$ (white squares) of $f^{-1}(\{\eta_2\})$ (black dots). Right: prediction set $S = f^{-1}(\{\eta_2\}) \cup \{z_\pm\}$ (white squares) of $f^{-1}(\{\eta_3\})$ (black dots).

Theorems 3.2, 3.4 and 3.6 establish the transport of images method if the points η_k are chosen properly. This motivates the next definition.

Definition 3.9. Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$. We call $(\eta_1, \dots, \eta_n) \in \mathbb{C}^n$ a *transport path* from η_1 to η_n if for each $k = 1, \dots, n-1$ either $f^{-1}(\{\eta_k\})$ or S in (3.7) is a prediction set of $f^{-1}(\{\eta_{k+1}\})$.

Along a transport path, all solutions of $f(z) = \eta_{k+1}$ can be computed from the solutions of $f(z) = \eta_k$. Next we show the existence of such paths. Here we call η a multiple caustic point if $|f^{-1}(\{\eta\}) \cap \mathcal{C}| > 1$.

Theorem 3.10. *Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$ without non-isolated multiple caustic points, and let η_s, $\eta_e \in \mathbb{C} \setminus f(\mathcal{C})$. Then there exists a transport path from η_s to η_e.*

Proof. We show that there exists a polygonal path from η_s to η_e, which intersects the caustics only at a finite number of simple fold points. Then this path is refined if necessary.

Let R be the complex plane $\mathbb{C}$ without caustic points that are not simple fold points, which is $\mathbb{C}$ minus a set of isolated points, and hence a region (open and connected), as we discuss next. Since f is non-degenerate, the functions $\partial_z f$, $\overline{\partial_{\bar{z}} f}$ and ω are rational, see (2.16). This has several implications. The sets $\mathcal{M}$ from (2.7) and $f(\mathcal{M})$ are finite. The set $\mathcal{C} \setminus \mathcal{M}$ consists of $\deg(\omega)$ many pre-images of the unit circle under ω. There are at most finitely many critical points z_0 of f with $\omega'(z_0) = 0$, hence the number of caustic points where the tangent (2.11) does not exist is finite. The number of caustic points where the tangent exists and is zero is finite by Lemma 2.3 (on each of the $\deg(\omega)$ many pre-images of the unit circle, there are at most finitely many points with $\tau = 0$, or the whole arc is mapped onto a single point). By assumption, multiple caustic points are isolated.

Next, we prove the existence of a rectifiable path from η_s to η_e in R with only finitely many caustic points (simple folds) on it. Since $\eta_s, \eta_e \notin f(\mathcal{C})$

and since the caustics are compact there exists $\delta > 0$ such that $D_\delta(\eta_s)$, $D_\delta(\eta_e)$ contain no caustic points. We consider the increasing sequence of regions $U_k \subseteq R$ whose points are 'at most k caustic crossings distant' from η_s, constructed as follows. Let U_0 be the component (maximal open and connected set) of $R \setminus f(\mathcal{C})$ with $\eta_s \in U_0$. Suppose that $U_k \subseteq R$ has been constructed. Let U_{k+1} be the component of $R \setminus (f(\mathcal{C}) \setminus (\partial U_0 \cup \ldots \cup \partial U_k))$ with $\eta_s \in U_{k+1}$. If $\eta_e \in U_k$, there exists a rectifiable path from η_s to η_e in R that intersects $\partial U_0, \ldots, \partial U_{k-1}$, and hence has k caustic crossings. If $\eta_e \notin U_k$ then the boundary of U_k in R, which consists of caustic arcs, has length at least $2\pi\delta$, since then $D_\delta(\eta_s) \subseteq U_k$ and $D_\delta(\eta_e) \subseteq R \setminus U_k$. Let $m = \lceil \frac{L}{2\pi\delta} \rceil$, where $L < \infty$ is the total length of the caustics (see the discussion below Definition 2.5). Then $\eta_e \in U_m$, and there exists a rectifiable path from η_s to η_e with at most m caustic points.

By manipulations in an arbitrary small neighborhood around this path in R we obtain a polygonal path $P = (\eta_1, \ldots, \eta_n)$ with $\eta_1 = \eta_s$ and $\eta_n = \eta_e$, such that: (1) $\eta_1, \ldots, \eta_n \in \mathbb{C} \setminus f(\mathcal{C})$, (2) each line segment $[\eta_k, \eta_{k+1}]$ contains at most one caustic point and (3) if the line segment $[\eta_k, \eta_{k+1}]$ contains a caustic point then $\arg(\eta_{k+1} - \eta_k) = \arg(\pm c)$, with c as in Theorem 3.5.

We refine this path to get a transport path. First we consider the line segments $[\eta_k, \eta_{k+1}]$, which contain a caustic point η, but where S in (3.7) is not a prediction set of $f^{-1}(\{\eta_{k+1}\})$. We add two points from $[\eta_k, \eta_{k+1}] \setminus f(\mathcal{C})$ sufficiently close to η such that Theorem 3.6 applies to these points. Denote the refined path for simplicity again by $P = (\eta_1, \ldots, \eta_n)$. Next we refine the line segments $E = [\eta_k, \eta_{k+1}]$ without caustic points. For all $\eta \in E$ there exists an $\varepsilon(\eta) > 0$ such that $f^{-1}(\{\eta\})$ is a prediction set of $f^{-1}(\{\xi\})$ for all $\xi \in D_{\varepsilon(\eta)}(\eta)$ by Theorem 3.4. The family $(D_{\varepsilon(\eta)}(\eta))_{\eta \in E}$ is an open covering of the compact set E. Hence, there exists a finite subcovering $(D_{\varepsilon(\kappa_j)}(\kappa_j))_{j=1,\ldots,\ell}$ with $\kappa_1 = \eta_k$ and $\kappa_\ell = \eta_{k+1}$, which gives a partition of E, where $f^{-1}(\{\kappa_j\})$ is a prediction set of $f^{-1}(\{\kappa_{j+1}\})$ for $j = 1, \ldots, \ell - 1$. Refining all line segments without caustic points yields a transport path. □

Note that non-degenerate harmonic mappings can actually have non-isolated multiple caustic points.

Example 3.11. The harmonic mapping $f(z) = \frac{1}{2}(z^2-1)^2 + \overline{z^2-1}$ from [37, Ex. 5.1] is non-degenerate and maps its critical arcs $\gamma_\pm : [-\pi, \pi] \to \mathcal{C}$, $\gamma_\pm(t) = \pm\sqrt{1 + e^{-it}}$, onto the same caustic arc.

We close this section by noting the correctness of the transport of images method, provided that 0 is not a caustic point of f.

Corollary 3.12. *Let f be a non-degenerate harmonic mapping on $\widehat{\mathbb{C}}$ without non-isolated multiple caustic points, such that 0 is not a caustic point. Then there exists a point $\eta_1 \in \mathbb{C} \setminus f(\mathcal{C})$, such that Theorem 3.2 applies to all poles of f, and there exists a transport path from η_1 to 0.*

Proof. Theorem 3.2 applies to all poles of f for all $\eta_1 \in \mathbb{C}$ with large enough $|\eta_1|$. Since $f(\mathcal{C})$ is compact, η_1 can be chosen in $\mathbb{C} \setminus f(\mathcal{C})$. Then there exists a transport path from η_1 to 0 by Theorem 3.10. □

Remark 3.13. In Corollary 3.12, the point 0 is not a caustic point of f. If 0 is a caustic point then f has (at least) one singular zero $z_0 \in \mathcal{C}$. Since continuation may run into numerical trouble if solutions are almost singular, special strategies are used. These are usually referred to as the *endgame*. For systems of analytic functions, two commonly used endgame strategies are the *Cauchy endgame* [27] and the *power series endgame* [28]. The setting for the transport of images method is somewhat different since we require the critical set $\mathcal{C}$ in the method. Therefore, we can (numerically) compute the zeros of f in $\mathcal{C}$, which are exactly the singular zeros. Thus, in the endgame we would only have to determine which homotopy curves intersect at z_0. This can be done with (2) and (3a) in Section 3.4 if f is non-degenerate and $\omega'(z_0) \neq 0$.

3.4 Analysis of the homotopy curves

We now consider the continuous problem behind our (discrete) computation and describe the homotopy curves, i.e., how the solution set of $f(z) = \eta(t)$ varies with t, where $f : \Omega \to \mathbb{C}$ is a harmonic mapping and $\eta : [a, b] \to \mathbb{C}$, $t \mapsto \eta(t)$, is a (continuous) path. We analyze the solution set locally with results from [23, 36, 37]. Combining the local results gives the global picture. We distinguish the cases (1) $z_0 \in \Omega \setminus \mathcal{C}$, (2) $z_0 \in \mathcal{M}$ and (3) $z_0 \in \mathcal{C} \setminus \mathcal{M}$.

(1) Let $z_0 \in \Omega \setminus \mathcal{C}$, i.e., the Jacobian of f is non-zero at z_0. By the inverse function theorem there exists an $\varepsilon > 0$ and an open set $V \subseteq \mathbb{C}$ such that $f : D_\varepsilon(z_0) \to V$ is a diffeomorphism. Thus, a curve $\eta(t)$ in V has a unique pre-image curve $z(t) = f^{-1}(\eta(t))$. In particular, the homotopy curves can intersect only at critical points.

(2) Let $z_0 \in \mathcal{M}$ from (2.7). Then f has the form

$$f(z) = a_0 + \sum_{k=n}^{\infty} a_k (z - z_0)^k + \overline{b_0 + \sum_{k=n}^{\infty} b_k (z - z_0)^k}$$

with $n \geq 2$ and $|a_n| \neq |b_n|$. If $\eta(t)$ passes through $f(z_0)$ then exactly n homotopy curves intersect at z_0 with equispaced angles, since $f(z) = \eta(t)$ is locally to leading order $f(z_0) + a_n(z - z_0)^n + \overline{b_n(z - z_0)^n} = \eta(t)$. The latter can be uniquely solved for $(z - z_0)^n$ by

$$(z - z_0)^n = \frac{\overline{a}_n(\eta(t) - f(z_0)) - \overline{b_n(\eta(t) - f(z_0))}}{|a_n|^2 - |b_n|^2},$$

see [36, Lem. 4.2], which yields the n solutions.

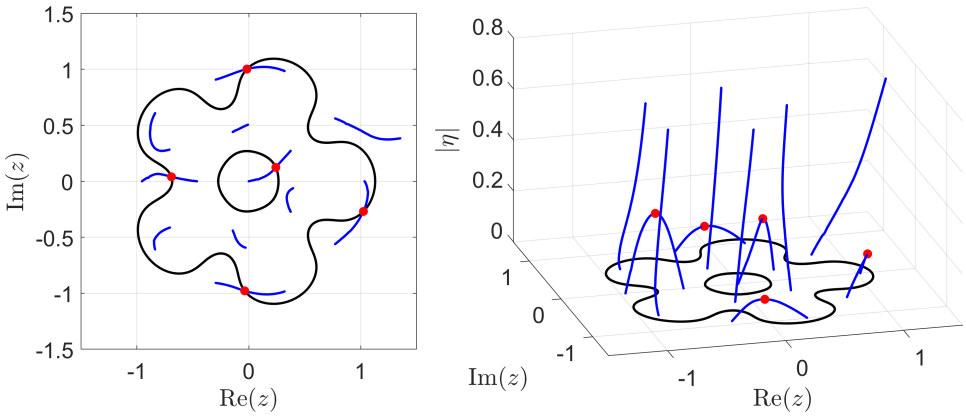


Figure 5: Computed homotopy curves (blue) with turning points (red dots) in Example 3.14 and critical set of f (black).

(3) For $z_0 \in \mathcal{C} \setminus \mathcal{M}$ the situation is more involved. By [23, Thm. 2.1], a harmonic mapping is either (a) light, (b) has zero Jacobian or (c) is constant on some arc of $\mathcal{C} \setminus \mathcal{M}$. We discuss these three cases.

(3a) Let f be a light harmonic mapping. Essentially, the behavior of the homotopy curves at z_0 depends on the tangent to the caustics at $f(z_0)$ and hence on ψ from (2.12). Let $z_0 \in \mathcal{C} \setminus \mathcal{M}$. Then z_0 is a zero of order $\ell \geq 0$ of $\partial_z f$ (and of $\overline{\partial_{\bar{z}} f}$). First let $\omega'(z_0) \neq 0$, so that $z_0 = \gamma(t_0)$ is a point on a critical curve that is not a self-intersection point of the critical curves.

1. If $\psi(t_0)$ is non-zero ($f(z_0)$ is a fold caustic point and consequently $\ell = 0$) then z_0 is a turning point of the homotopy curves, i.e., two curves start or end at z_0 if $\eta(t)$ crosses the caustics at $f(z_0)$. This follows from Theorem 3.5 or [23, Thm. 5.1] and is illustrated in [37, Fig. 5].

2. More generally, if ψ does not change sign at t_0 then [23, Thm. 5.1] implies that ℓ homotopy curves intersect at z_0 and two homotopy curves start or end at z_0, if $\eta(t)$ crosses the caustics at $f(z_0)$. Hence, z_0 is a turning point and, if $\ell > 0$, a bifurcation point of the homotopy curves.

3. If ψ changes sign at t_0 (i.e., $f(z_0)$ is a cusp) then by [23, Thm. 5.1] z_0 is a bifurcation point where $\ell + 1$ homotopy curves intersect and two homotopy curves start or end, if $\eta(t)$ crosses the caustics at $f(z_0)$.

For $\omega'(z_0) = 0$ there is only an upper bound on the local valence of f at z_0 in [23, Sect. 6] and hence on the number of homotopy curves, which also depends on the caustic tiles bordering $f(z_0)$ that are traversed by $\eta(t)$.

Example 3.14. Consider the harmonic mapping (5.2) with $n = 5$ and $r = 0.6$, which is light and non-degenerate. Figure 5 shows the computed homotopy curves where $\eta = \eta(t)$ parametrizes the segment from

$0.6699 + 0.1795i$ to 0, which intersects the caustics of f only at simple fold points. Along this path each caustic crossing produces two additional solutions. The corresponding turning points of the homotopy curves are marked with red dots. Figure 5 shows the solutions $z(t)$ of $f(z) = \eta(t)$ in the z-plane (left) and over the z-plane at height $|\eta(t)|$ (right).

(3b) Let f be a harmonic mapping with $J_f \equiv 0$, i.e., $\mathcal{C} = \Omega$. Then f has the form $f(z) = a\,\mathrm{Re}(h(z)) + b$ with $a, b \in \mathbb{C}$, $a \neq 0$, and an analytic function h; see [23, Lem. 2.1] or [30, Lem. 4.7]. (If Ω is not simply connected, h may be multi-valued, but $\mathrm{Re}(h)$ is single-valued.) Then $f(\Omega) \subseteq \{at + b : t \in \mathbb{R}\}$, and $f(z) = \eta$ if and only if $\mathrm{Re}(h(z)) = (\eta - b)/a$. Since h is analytic, it is either constant or has the open mapping property, which shows that the solution set of $f(z) = \eta$ is either empty, Ω or a union of curves. Therefore, along a path $\eta(t)$ crossing the caustics of f, the solutions do not form curves depending on t.

Example 3.15. The Jacobian of $f(z) = z + \overline{z}$ vanishes identically. Clearly, $f(z) = \eta \in \mathbb{C} \setminus \mathbb{R}$ has no solution, while $f(z) = \eta \in \mathbb{R}$ is solved by all z with $\mathrm{Re}(z) = \eta/2$.

(3c) Let f be a harmonic mapping that is constant on some arc Γ of $\mathcal{C} \setminus \mathcal{M}$, and $z_0 \in \Gamma$. Then f maps Γ onto the single point $f(z_0)$. For $\eta(t)$ not equal to $f(z_0)$ (or any similar point), the homotopy curves are as for light harmonic mappings. However, if $\eta(t) = f(z_0)$ then the solution set also contains all points of Γ, and the solutions do not form curves depending on t.

Example 3.16. The non-degenerate harmonic mapping $f(z) = z - 1/\overline{z}$ is not light since $f^{-1}(\{0\}) = \{z \in \mathbb{C} : |z| = 1\} = \mathcal{C}$. Let $\eta \in \mathbb{C} \setminus \{0\}$. Then $f(z) = \eta$ has the two solutions

$$z_\pm(\eta) = \frac{1 \pm \sqrt{1 + 4|\eta|^{-2}}}{2}\eta$$

which satisfy $|z_+(\eta)| > 1$ and $|z_-(\eta)| < 1$, and $|z_\pm(\eta)| \to 1$ if $\eta \to 0$. Figure 6 shows a path $\eta(t)$ crossing the origin (panel 1) and the corresponding homotopy curves (panel 2). For $\eta(t) = 0$ the two solutions are replaced by all points on the unit circle. Panels 3 and 4 show a similar example with a non-smooth path $\eta(t)$, which leads to homotopy curves exiting $\mathcal{C}$ at different points.

The analysis of the homotopy curves also confirms that Definition 3.9 of a transport path is reasonable. The points η_k are in $\mathbb{C} \setminus f(\mathcal{C})$ to avoid a zero Jacobian and potentially infinitely many solutions. Moreover, to avoid bifurcation points of the homotopy curves, transport paths cross the caustics only at fold points, which can be handled by Theorem 3.5.

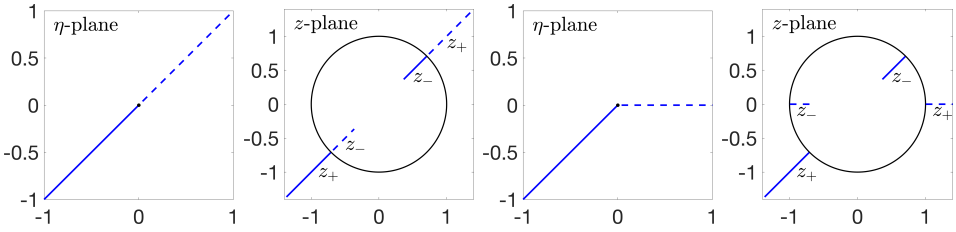


Figure 6: Homotopy curves of $f(z) = z - 1/\overline{z}$ for two paths $\eta(t)$ (panels 1 and 3), divided in a solid and dashed blue part, and corresponding pre-images (panels 2 and 4); see Example 3.16. The black circle is $\mathcal{C} = f^{-1}(\{0\})$.

4 Implementation

A MATLAB implementation of our method is freely available at

`https://github.com/transportofimages/`

This toolbox contains routines to compute the zeros of non-degenerate harmonic mappings and the critical curves and caustics. Moreover, it contains m-files to reproduce all examples in Section 5. We briefly describe two key aspects of our implementation.

4.1 Computation of the caustics

The transport of images methods bases essentially upon the correct handling of caustic crossings. To compute the caustics we first compute the critical curves. Critical points $z \in \mathcal{C} \setminus \mathcal{M}$ satisfy $\omega(z) = e^{it}$ with $t \in [0, 2\pi[$; see (2.9). Since ω is rational for non-degenerate harmonic mappings, $\omega(z) = e^{it}$ is equivalent to a polynomial equation with $\deg(\omega)$ many solutions. We first solve $\omega(z) = 1$ via the polynomial equation. Then, for $0 = t_1 < t_2 < \ldots < t_k < 2\pi$ we successively solve $\omega(z) = e^{it_{j+1}}$ with Newton's method (with initial points $\omega^{-1}(\{e^{it_j}\})$). If successful, this procedure gives $\deg(\omega)$ many (discretized) critical arcs, parametrized according to (2.10). Gluing them together yields the critical curves. Finally, the image under f of the critical curves is a discretization of the caustics of f.

4.2 Construction of transport paths

We construct a transport path to 0 along $R_\theta = \{re^{i\theta} : r > 0\}$, $\theta \in [0, 2\pi[$. Beneficially, the intersection points of the caustic and the ray R_θ can be read off from the argument of the caustic points. Also, as discussed in Remark 3.7, crossing the caustics in a specific direction is not essential, as long as the crossing direction is not tangential to the caustics. For brevity we still call $(\eta_1, \ldots, \eta_n) \in \mathbb{C}^n$ a transport path if $f^{-1}(\{\eta_k\})$ or S in (3.7) is a prediction set of $f^{-1}(\{\eta_{k+1}\})$, even if the step from η_k to η_{k+1} is not in the direction c in

Theorem 3.6, i.e., even if $\arg(\eta_{k+1} - \eta_k) \neq \arg(\pm c)$. Determining a suitable angle $\theta \in [0, 2\pi[$ for R_θ can be challenging. In particular, the transport phase (along R_θ) may fail if too small step sizes are required, or if R_θ is tangential or almost tangential to the caustics, or if R_θ intersects the caustic at or near cusps or other non-fold points. To overcome this difficulty we choose the angle $\theta \in [0, 2\pi[$ at random (e.g., uniformly distributed). If the transport phase with θ is not successful we restart with a new random angle.

We construct a transport path to 0 along R_θ as follows:

(1) Set $\eta_1 = 2e^{i\theta} \max_{z\in\mathcal{C}} |f(z)|$. If Theorem 3.2 with η_1 does not apply to all poles of f, e.g., two distinct points ζ_j and ζ_ℓ are attracted by the same solution of $f(z) = \eta_1$, we increase $|\eta_1|$ and proceed. Eventually, $|\eta_1|$ is large enough and Theorem 3.2 applies to all poles of f simultaneously.

(2) Let $\{\xi_1, \ldots, \xi_\ell\} = f(\mathcal{C}) \cap R_\theta$ with $|\xi_1| > \ldots > |\xi_\ell| > 0$. We compute these intersection points from the discretized caustics with bisection. For each ξ_k let $\eta_{2k}, \eta_{2k+1} \in R_\theta$ with equal distance to ξ_k, and such that $|\eta_1| > |\eta_2| > \ldots > |\eta_{2\ell+1}| > 0$. Then $P = (\eta_1, \eta_2, \ldots, \eta_{2\ell+1}, 0)$ is a potential transport path.

(3) We refine P recursively with the following divide-and-conquer scheme to obtain a transport path.

(3a) If the step from η_j to η_{j+1}, where the line segment $[\eta_j, \eta_{j+1}]$ does not intersect the caustics, is unsuccessful, i.e., if $f^{-1}(\{\eta_j\})$ is not a prediction set of $f^{-1}(\{\eta_{j+1}\})$, we divide the step from η_j to η_{j+1} into two substeps by inserting $\eta' = (\eta_j + \eta_{j+1})/2$ into P.

(3b) If the step from η_j to η_{j+1}, where the line segment $[\eta_j, \eta_{j+1}]$ intersects the caustic, is unsuccessful, i.e., if S in (3.7) is not a prediction set of $f^{-1}(\{\eta_{j+1}\})$, we divide the step from η_j to η_{j+1} into three substeps by inserting $\eta_1' = (3\eta_j + \eta_{j+1})/4$ and $\eta_2' = (\eta_j + 3\eta_{j+1})/4$ into P.

(4) We proceed, dividing the steps in (3a) and (3b) until P is a transport path, or until we reached a prescribed recursion depth. In the latter case we restart with a new random angle θ in (1).

5 Numerical examples

In order to verify the numerical results we test our method on the following functions, for which the number of zeros is known analytically.

1. Wilmshurst's harmonic polynomial

$$f(z) = (z-1)^n + z^n + \overline{i(z-1)^n - iz^n}, \quad n \geq 1, \tag{5.1}$$

is non-degenerate and has exactly n^2 zeros. This is the maximum number of zeros of a harmonic polynomial $f(z) = p(z) + \overline{q(z)}$ with $\deg(p) = n$ and $\deg(q) < n$; see [44].

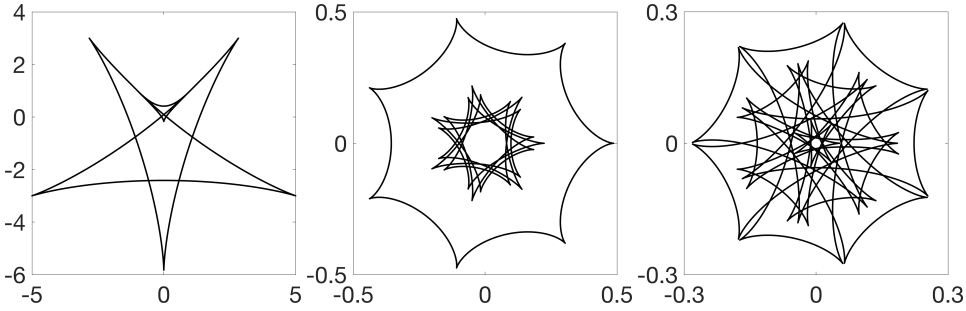


Figure 7: Caustics of harmonic mappings: (5.1) with $n = 3$ (left), (5.2) with $n = 7$ and $\rho = 0.7$ (middle), (5.3) with $n = 7$, $\rho = 0.7$ and $\varepsilon = 0.1$ (right).

2. The rational harmonic mapping of Mao, Petters and Witt [24],

$$f(z) = z - \overline{\left(\frac{z^{n-1}}{z^n - \rho^n}\right)}, \quad n \geq 3, \quad \rho > 0, \tag{5.2}$$

is non-degenerate and has exactly $3n+1$ zeros for $\rho < \rho_c$, $2n+1$ zeros for $\rho = \rho_c$ and $n+1$ zeros for $\rho > \rho_c$, where

$$\rho_c = \left(\frac{n-2}{n}\right)^{\frac{1}{2}} \left(\frac{2}{n-2}\right)^{\frac{1}{n}};$$

see [21, Prop. 2.1].

3. Rhie's rational harmonic mapping

$$f(z) = z - \overline{\left((1-\varepsilon)\frac{z^{n-1}}{z^n - \rho^n} + \frac{\varepsilon}{z}\right)}, \quad n \geq 3, \quad \rho, \varepsilon > 0, \tag{5.3}$$

is non-degenerate and has exactly $5n$ zeros if $\rho < \rho_c$ (as above) and if $\varepsilon < \varepsilon^*$, where ε^* depends on n and ρ; see [21, Thm. 2.2] for details. This is the maximum number of zeros of a harmonic mapping $f(z) = z - \overline{r(z)}$ with a rational function r of degree $n+1 \geq 2$; see [14, Thm. 1].

These functions should be challenging for the transport of images method since they have a large number of zeros and very nested caustics; see Figure 7. In gravitational lensing, the functions (5.2), (5.3) and their zeros are of interest; see [15, 21].

The following computations have been performed in MATLAB R2019b on an i7-7700 4×3.60 GHz CPU with 16 GB RAM using our implementation[3], which also contains m-files to reproduce all examples.

First, we compute the zeros of (5.1), (5.2), (5.3), and of the transcendental harmonic mapping (2.19) with the transport of images method. Here

[3] `https://github.com/transportofimages/`

Function	Zeros	Max. res.	Time	N. iter.	Steps	Ref.
(2.19)	4	6.6613e-16	36 ms	237	8	2
(5.1)	9	1.7764e-15	44 ms	291	8	1
(5.2)	22	8.9509e-16	62 ms	1646	19	4
(5.3)	35	7.8505e-16	85 ms	3851	40	9

Table 1: The transport of images method for (2.19), (5.1) with $n = 3$, (5.2) with $n = 7$, $\rho = 0.7$, and (5.3) with $n = 7$, $\rho = 0.7$, $\varepsilon = 0.1$: maximal residual (Max. res.), computation time, number of harmonic Newton iterations (N. iter.), number of transport steps, and number of step refinements (Ref.).

we fix the angle $\theta = \pi/50$ in the construction of the transport path (see Section 4.2). The results are displayed in Table 1. We make the following three main observations: (1) The transport of images method finishes with the correct number of zeros for each function, in particular for the transcendental harmonic mapping (2.19). The logarithmic term makes a symbolic computation of the zeros difficult, e.g., Mathematica cannot determine the zeros of this function. We emphasize that the number of zeros is not an input parameter of our method. (2) The results are computed within a fraction of a second on a standard computer. (3) In each case the computed results are very accurate, as shown by the residual $|f(z_j)|$ at the computed zeros, which is in the order of the machine precision. Table 1 shows the maximal residual over all computed zeros for the respective functions. Furthermore, we see that a step refinement as described in Section 4.2 was necessary for all functions. The number of (harmonic) Newton iterations also includes iterations that are necessary to decide whether a step has to be refined.

Figure 8 illustrates the transport of images method for (5.1) with $n = 3$. Panels 1 and 2 display the transport path and a zoom-in close to the origin. The remaining panels show the dynamics of the harmonic Newton maps $H_{f-\eta_k}$ for $k = 1, 2, \ldots, 9$. Recall from Remark 3.8 that points which are attracted by the same solution have the same color. The initial phase is visualized in the top right plot. Every solution of $f(z) = \eta_1$ (black dots) attracts exactly one of the initial points (white squares), constructed as in Section 3.2, hence these initial points form a prediction set of $f^{-1}(\{\eta_1\})$. For aesthetic reasons we have set $\eta_1 = 6e^{i\pi/50}$ by hand. The remaining panels visualize the steps in the transport phase, again with the solutions of $f(z) = \eta_k$ (black dots) and the prediction sets (white squares). In particular, we see how the basins of attraction evolve while η_k 'travels' along the transport path. New basins and their respective solutions appear in pairs for η_3 (green and cyan), η_5 (blue and light blue) and η_7 (pink and purple), right after a caustic crossing, as predicted by the theory. For each step the prediction set of $f^{-1}(\{\eta_{k+1}\})$ consists of $f^{-1}(\{\eta_k\})$ (black dots in the previous panel), and of the points $z_\pm$ from Theorem 3.5 in case of a caustic crossing. We

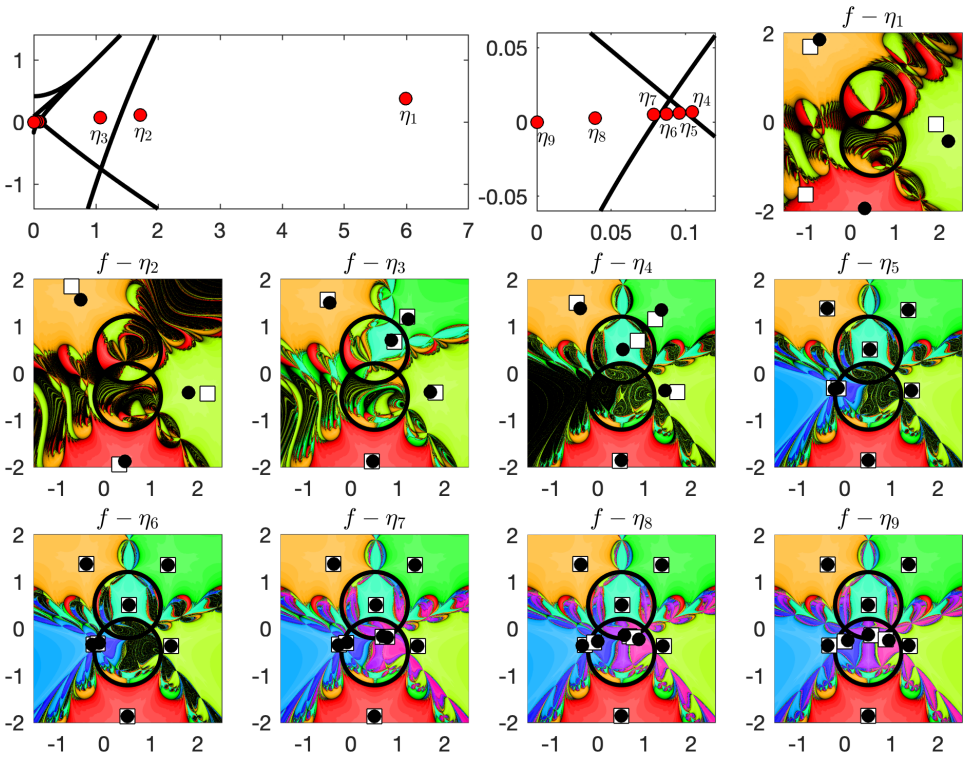


Figure 8: Transport of images method for (5.1) with $n = 3$: Transport path and dynamics of $H_{f-\eta_k}$ with prediction set (white squares) and zeros of $f-\eta_k$ (black dots).

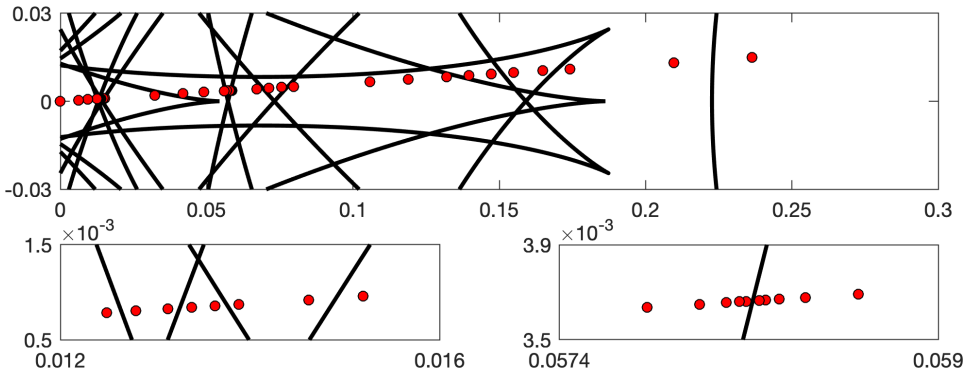


Figure 9: Transport path for (5.3) with $n = 7$, $\rho = 0.7$, $\varepsilon = 0.1$ and $\theta = \frac{\pi}{50}$.

see that the 'new' solutions are close to the predicted points $z_\pm$. Note that Theorem 3.4 does not apply to the step from η_7 to η_9. Therefore, this step was divided by introducing the point $\eta_8 = (\eta_7+\eta_9)/2$. A transport path with more step refinements, but for the function (5.3), is displayed in Figure 9.

Next, we compute the zeros of the function (5.2) for different values

n	Zer.	N. iter.	Rest.	Time
3	10	543.98	0.00	37 ms
4	13	926.68	0.00	45 ms
5	16	1328.06	0.02	50 ms
6	19	1432.50	0.00	53 ms
7	22	1765.74	0.02	60 ms
8	25	2253.86	0.02	68 ms
9	28	2653.40	0.00	73 ms
10	31	3126.28	0.04	81 ms
11	34	3393.04	0.00	88 ms
12	37	3926.20	0.02	97 ms

n	Zer.	N. iter.	Rest.	Time
3	15	948.46	0.00	44 ms
4	20	1613.32	0.02	55 ms
5	25	2046.24	0.02	61 ms
6	30	3014.40	0.00	70 ms
7	35	3590.90	0.04	81 ms
8	40	4989.04	0.00	93 ms
9	45	5820.60	0.06	107 ms
10	50	6941.06	0.06	120 ms
11	55	8035.50	0.10	137 ms
12	60	9154.42	0.04	147 ms

Table 2: The transport of images method averaged over 50 instances. Left: (5.2) for $\rho \in [0.7, \rho_c[$. Right: (5.3) for $\rho \in [0.7, \rho_c[$ and $\varepsilon \in [\frac{\varepsilon^*}{2}, \frac{3\varepsilon^*}{4}]$. Number of harmonic Newton iterations (N. iter.), number of restarts with new angle θ (Rest.), computation time.

Function	Zeros	Max. res.	Restarts	Time
(5.2)	301	6.4737e-16	0	5.36 s
(5.3)	125	1.3552e-15	7	2.52 s

Table 3: The transport of images method for (5.2) with $n = 100$, $\rho = 0.94$, and (5.3) with $n = 25$, $\rho = 0.9$, $\varepsilon = 0.4$.

of n with the transport of images method (without fixing the angle θ). For each $n = 3, 4, \ldots, 12$ we consider 50 instances of (5.2) with randomized $\rho \in [0.7, \rho_c[$, and apply our method to these functions. Similarly, we consider 50 instances of the function (5.3) with random $\rho \in [0.7, \rho_c[$ and $\varepsilon \in [\frac{\varepsilon^*}{2}, \frac{3\varepsilon^*}{4}]$. The quantities ρ and ε are uniformly distributed in the respective intervals. Table 2 shows the computed number of zeros, the number of harmonic Newton iterations, the number of restarts with a new angle θ and the computation time. For each n, the quantities are averaged over all instances. We observe the following. First of all, the transport of images method terminates with the correct number of zeros for all n and all instances. For most instances the transport along the first ray is already successful. For the remaining ones a restart with a new angle is necessary to compute the zeros. In this example, computing the caustics takes roughly half of the overall computation time. This can be exploited when solving $f(z) = \eta_k$ for several $\eta_k \in \mathbb{C}$, which is a scenario in the astrophysical application in [33, Sect. 10.5].

In our next example we consider (5.2) with $n = 100$ and $\rho = 0.94$, and (5.3) with $n = 25$, $\rho = 0.9$ and $\varepsilon = 0.4$. Note that every transport path along a ray R_θ intersects the caustics at least 100 times for the first function and at least 49 times for the second function. The transport of images method computes all zeros in a few seconds on a standard computer; see

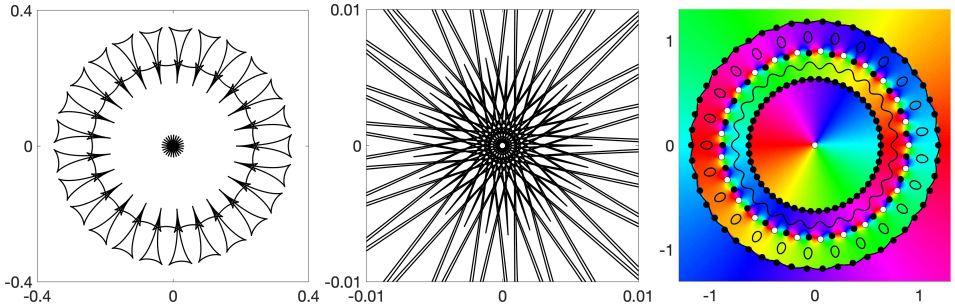


Figure 10: The harmonic mapping $f(z) = z - \overline{\left(0.6\frac{z^{24}}{z^{25}-0.9^{25}} + 0.4\frac{1}{z}\right)}$; see (5.3). Left: caustics. Middle: zoom of caustics. Right: phase plot with critical curves, zeros (black dots) and poles (white dots).

Table 3. The residuals are close to the machine precision. In contrast to the examples in Table 2, seven restarts with a new angle θ are necessary for (5.3) with $n = 25$. This is due to the extremely nested caustics; see Figure 10 (left and middle). The 125 zeros and 26 poles of the second function are displayed in a phase plot in Figure 10. In a phase plot of f the complex plane is colored according to the phase $f(z)/|f(z)|$; see [43] for an extensive discussion of phase plots and their applications.

As mentioned in the introduction we are not aware of any specialized software to compute *all* zeros of harmonic mappings. However, we compare our method with the general purpose root finder from [29], which is based on Bézout resultants. There are two implementations of this method: `rootsb`[4] and the `roots` command of Chebfun2[5]. Chebfun2 [42] is the state-of-the-art toolbox for numerical computation with (real or complex) smooth and bounded functions on a rectangle in the plane. The idea of [29] is to find the zeros of two (real-valued) smooth functions $f(x, y) = g(x, y) = 0$ by first computing polynomial approximants $p(x, y)$ and $q(x, y)$ that are accurate to machine precision in the supremum norm relatively to f and g, respectively, and then solving the polynomial system $p(x, y) = q(x, y) = 0$. In this procedure the polynomial approximants are locally resampled if appropriate. While `rootsb` uses the original functions, Chebfun2 uses the computed interpolant for the resampling.

We consider again the functions in Table 2. To compute the zeros of (5.2) and (5.3) with Chebfun2 and `rootsb` we multiply them by their respective denominators to remove the poles and thus obtain the functions

$$F_1(z) = z(\overline{z}^n - \rho^n) - \overline{z}^{n-1},$$
$$F_2(z) = (|z|^2 - 1)\overline{z}^n + (\varepsilon - |z|^2)\rho^n.$$

[4]rootsb, version of March 20, 2021, `https://de.mathworks.com/matlabcentral/fileexchange/44084-computing-common-zeros-of-two-bivariate-functions`.

[5]Chebfun2, version of September 30, 2020, `www.chebfun.org`.

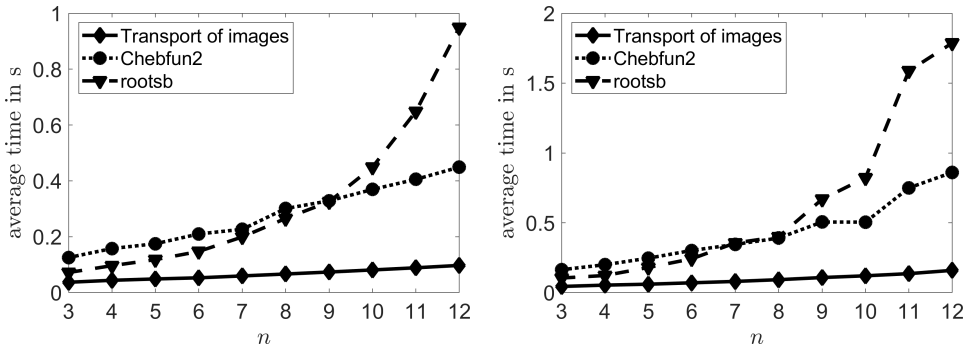


Figure 11: Timings of the transport of images method, Chebfun2 and rootsb averaged over the 50 instances for (5.2) (left) and (5.3) (right) as in Table 2.

Function	Transport of images		Chebfun2		rootsb	
	Zeros	Time	Zeros	Time	Zeros	Time
(5.2)	301	5.36 s	10017	45.73 s	301	60.93 s
(5.3)	125	2.52 s	95	9.98 s	75	10.78 s

Table 4: Comparison of the transport of images method with Chebfun2 and `rootsb` for the functions (5.2) with $n = 100$, $\rho = 0.94$, and (5.3) with $n = 25$, $\rho = 0.9$, $\varepsilon = 0.4$, see also Table 3.

We use the square $R = [-1.2, 1.2] \times [-1.2, 1.2]$, which contains all zeros of F_1 and F_2 for all $n = 3, 4, \ldots, 12$ and all ρ and ε as above. Note that the supremum norm of F_j on R grows exponentially with n. Despite this difficulty, Chebfun2 and `rootsb` compute all zeros of F_1 for all instances. For F_2, `rootsb` computes all zeros of all instances, and Chebfun2 returns the correct number of zeros for all instances for $n = 3, 4, \ldots, 9$ and $n = 11$, but a wrong number of zeros for 5 out of 50 instances for $n = 10$ or $n = 12$. This may be due to the fact that the magnitude of F_2 (and also F_1) varies highly, which is known as the 'dynamical range issue'. Here, resampling with the original function in `rootsb` leads to more accurate results at the expense of speed; see Figure 11. However, for larger n as in Table 4, `rootsb` also does not compute the correct zeros. This highlights the ill-behavedness of the considered functions and how difficult it is to compute their zeros. The problem-adapted transport of images method finishes in all cases with the correct number of zeros; see Table 2. Moreover, it is much faster than Chebfun2 and `rootsb` in these examples; see Figure 11 and Table 4.

6 Summary and outlook

We established and analyzed the transport of images method, which is the first problem-adapted method to compute *all* zeros of a (non-degenerate)

harmonic mapping f. Our method is guaranteed to find all zeros of f, even though this number is a priori unknown, provided that no zero of f is singular. Moreover, our MATLAB implementation performs remarkably well for several functions from the literature. While we focused on the computation of zeros in our examples, we can also compute all solutions of $f(z) = \eta$ for any point $\eta \in \mathbb{C} \setminus f(\mathcal{C})$. For this, we either consider $f - \eta$, or replace the transport path from η_1 to 0 by a transport path from η_1 to η (with η_1 from the initial phase). More generally, we can compute all solutions of $f(z) = \eta$ given all solutions of $f(z) = \xi$ (with any points $\eta, \xi \in \mathbb{C} \setminus f(\mathcal{C})$) and the caustics of f, by transporting the solutions along a transport path from ξ to η. This allows us to deduce the solutions for all right-hand sides from the local behavior of f on the critical curves.

Further studies of the transport of images method could include different constructions of transport paths (e.g., step size control), improvements in the computation of the caustics, as well as the investigation of other algorithms as corrector. Generalizing the transport of images method to a broader class of functions, e.g., for harmonic mappings on domains with boundary (see, e.g., [4]), might also be of interest. Finally, adjusting the transport of images method to the special functions of gravitational lensing should be of particular interest.

Acknowledgements We thank Jörg Liesen for several helpful suggestions on the manuscript. Moreover, we are grateful to the anonymous referees for valuable comments, which lead to improvements of this work.

References

[1] E. L. Allgower and K. Georg, *Introduction to numerical continuation methods*, vol. 45 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2003.

[2] J. Arango, H. Arbeláez, and J. Rivera, *Orientation at singularities of harmonic functions*, Monatsh. Math., 193 (2020), pp. 737–762.

[3] C. Bénéteau and N. Hudson, *A survey on the maximal number of solutions of equations related to gravitational lensing*, in Complex analysis and dynamical systems, Trends Math., Birkhäuser/Springer, Cham, 2018, pp. 23–38.

[4] W. Bergweiler and A. Eremenko, *On the number of solutions of a transcendental equation arising in the theory of gravitational lensing*, Comput. Methods Funct. Theory, 10 (2010), pp. 303–324.

[5] P. M. Bleher, Y. Homma, L. L. Ji, and R. K. W. Roeder, *Counting zeros of harmonic rational functions and its application to gravitational lensing*, Int. Math. Res. Not. IMRN, 2014 (2014), pp. 2245–2264.

[6] P. Duren, *Harmonic mappings in the plane*, vol. 156 of Cambridge Tracts in Mathematics, Cambridge University Press, Cambridge, 2004.

[7] P. Duren, W. Hengartner, and R. S. Laugesen, *The argument principle for harmonic functions*, Amer. Math. Monthly, 103 (1996), pp. 411–415.

[8] C. D. Fassnacht, C. R. Keeton, and D. Khavinson, *Gravitational lensing by elliptical galaxies, and the Schwarz function*, in Analysis and mathematical physics, Trends Math., Birkhäuser, Basel, 2009, pp. 115–129.

[9] L. Geyer, *Sharp bounds for the valence of certain harmonic polynomials*, Proc. Amer. Math. Soc., 136 (2008), pp. 549–555.

[10] W. Hengartner and G. Schober, *Univalent harmonic functions*, Trans. Amer. Math. Soc., 299 (1987), pp. 1–31.

[11] P. Henrici, *Applied and computational complex analysis. Vol. 3*, Pure and Applied Mathematics (New York), John Wiley & Sons, Inc., New York, 1986.

[12] K. L. Judd, *Numerical methods in economics*, MIT Press, Cambridge, MA, 1998.

[13] D. Khavinson and E. Lundberg, *Transcendental harmonic mappings and gravitational lensing by isothermal galaxies*, Complex Anal. Oper. Theory, 4 (2010), pp. 515–524.

[14] D. Khavinson and G. Neumann, *On the number of zeros of certain rational harmonic functions*, Proc. Amer. Math. Soc., 134 (2006), pp. 1077–1085.

[15] D. Khavinson and G. Neumann, *From the fundamental theorem of algebra to astrophysics: a "harmonious" path*, Notices Amer. Math. Soc., 55 (2008), pp. 666–675.

[16] D. Khavinson and G. Świątek, *On the number of zeros of certain harmonic polynomials*, Proc. Amer. Math. Soc., 131 (2003), pp. 409–414.

[17] S.-Y. Lee, A. Lerario, and E. Lundberg, *Remarks on Wilmshurst's theorem*, Indiana Univ. Math. J., 64 (2015), pp. 1153–1167.

[18] J. Liesen and J. Zur, *How constant shifts affect the zeros of certain rational harmonic functions*, Comput. Methods Funct. Theory, 18 (2018), pp. 583–607.

[19] J. Liesen and J. Zur, *The maximum number of zeros of $r(z) - \overline{z}$ revisited*, Comput. Methods Funct. Theory, 18 (2018), pp. 463–472.

[20] R. Luce and O. Sète, *The index of singular zeros of harmonic mappings of anti-analytic degree one*, Complex Var. Elliptic Equ., 66 (2021), pp. 1–21.

[21] R. Luce, O. Sète, and J. Liesen, *Sharp parameter bounds for certain maximal point lenses*, Gen. Relativity Gravitation, 46, Art. No. 1736 (2014), pp. 1–16.

[22] R. Luce, O. Sète, and J. Liesen, *A note on the maximum number of zeros of $r(z) - \overline{z}$*, Comput. Methods Funct. Theory, 15 (2015), pp. 439–448.

[23] A. Lyzzaik, *Local properties of light harmonic mappings*, Canad. J. Math., 44 (1992), pp. 135–153.

[24] S. Mao, A. O. Petters, and H. J. Witt, *Properties of point mass lenses on a regular polygon and the problem of maximum number of images*, in The Eighth Marcel Grossmann Meeting, Part A, B (Jerusalem, 1997), World Sci. Publ., River Edge, NJ, 1999, pp. 1494–1496.

[25] D. Mehta, T. Chen, J. D. Hauenstein, and D. J. Wales, *Communication: Newton homotopies for sampling stationary points of potential energy landscapes*, J. Chem. Phys., 141 (2014), p. 121104.

[26] A. Morgan, *Solving polynomial systems using continuation for engineering and scientific problems*, vol. 57 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009.

[27] A. P. Morgan, A. J. Sommese, and C. W. Wampler, *Computing singular solutions to nonlinear analytic systems*, Numer. Math., 58 (1991), pp. 669–684.

[28] A. P. Morgan, A. J. Sommese, and C. W. Wampler, *A power series method for computing singular solutions to nonlinear analytic systems*, Numer. Math., 63 (1992), pp. 391–409.

[29] Y. Nakatsukasa, V. Noferini, and A. Townsend, *Computing the common zeros of two bivariate functions via Bézout resultants*, Numer. Math., 129 (2015), pp. 181–209.

[30] G. Neumann, *Valence of complex-valued planar harmonic functions*, Trans. Amer. Math. Soc., 357 (2005), pp. 3133–3167.

[31] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York-London, 1970.

[32] A. O. Petters, *Gravity's action on light*, Notices Amer. Math. Soc., 57 (2010), pp. 1392–1409.

[33] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses*, Springer-Verlag, Berlin - Heidelberg, 1992.

[34] O. Sète, R. Luce, and J. Liesen, *Creating images by adding masses to gravitational point lenses*, Gen. Relativity Gravitation, 47, Art. No. 42 (2015), pp. 1–8.

[35] O. Sète, R. Luce, and J. Liesen, *Perturbing rational harmonic functions by poles*, Comput. Methods Funct. Theory, 15 (2015), pp. 9–35.

[36] O. Sète and J. Zur, *A Newton method for harmonic mappings in the plane*, IMA J. Numer. Anal., 40 (2020), pp. 2777–2801.

[37] O. Sète and J. Zur, *Number and location of pre-images under harmonic mappings in the plane*, Ann. Fenn. Math., 46 (2021). Online first publication.

[38] T. Sheil-Small, *Complex Polynomials*, vol. 75 of Cambridge Studies in Advanced Mathematics, Cambridge University Press, Cambridge, 2002.

[39] S. Smale, *Newton's method estimates from data at one point*, in The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985), Springer, New York, 1986, pp. 185–196.

[40] A. J. Sommese and C. W. Wampler, II, *The numerical solution of systems of polynomials*, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2005.

[41] T. J. Suffridge and J. W. Thompson, *Local behavior of harmonic mappings*, Complex Variables Theory Appl., 41 (2000), pp. 63–80.

[42] A. Townsend and L. N. Trefethen, *An extension of Chebfun to two dimensions*, SIAM J. Sci. Comput., 35 (2013), pp. c495–c518.

[43] E. Wegert, *Visual complex functions. An introduction with phase portraits*, Birkhäuser/Springer Basel AG, Basel, 2012.

[44] A. S. Wilmshurst, *The valence of harmonic polynomials*, Proc. Amer. Math. Soc., 126 (1998), pp. 2077–2081.

Zusammenfassung

In dieser Dissertationsschrift betrachten wir die Nullstellen harmonischer Funktionen in der komplexen Zahlenebene. Unsere Untersuchungen sind ursprünglich durch den Gravitationslinseneffekt in der Astrophysik motiviert, wo Spezialfälle harmonischer Funktionen und ihrer Nullstellen eine wichtige Rolle spielen. Wir konzentrieren uns in dieser Arbeit jedoch auf viel allgemeinere Funktionen. Unsere Ergebnisse sind sowohl von rein analytischer als auch numerischer Natur. Für rationale harmonische Funktionen $f(z) = r(z) - \overline{z}$ und harmonische Polynome $f(z) = p(z) + \overline{q(z)}$ verbessern wir bekannte Resultate über die Anzahl der Nullstellen. Außerdem entwickeln wir das erste numerische Fortsetzungsverfahren, welches garantiert alle Nullstellen einer (nicht-singulären und nicht-degenerierten) harmonischen Funktion f berechnet. Dieses Verfahren funktioniert ohne jegliche Vorkenntnisse über die Anzahl oder die Position der Nullstellen. Unsere numerischen Experimente unterstreichen außerdem die hohe Geschwindigkeit und große Genauigkeit. Grundlage für unsere analytischen und numerischen Ergebnisse ist eine umfassende Untersuchung darüber, wie sich die Anzahl der Lösungen von $f(z) = \eta$ mit dem Parameter $\eta \in \mathbb{C}$ verändert. Zu diesem Zweck verwenden wir Hilfsmittel aus ganz verschiedenen Bereichen der Mathematik, wie z.B. das Argumentenprinzip für harmonische Funktionen oder Konvergenzergebnisse für das Newton-Verfahren in Banach-Räumen. Wie sich herausstellt, ist die Position von η bezüglich der Kaustiken von f entscheidend zur Bestimmung der Anzahl der Lösungen von $f(z) = \eta$. Wir formalisieren diese Beobachtung und vervollständigen sie zu einer allgemeinen Theorie. Diese ist von eigenständigem Interesse und kann genutzt werden, um das Verständnis von harmonischen Funktionen zukünftig weiter zu verbessern.