How constant shifts affect the zeros of certain rational harmonic functions

Jörg Liesen* Jan Zur*

May 16, 2018

Abstract

We study the effect of constant shifts on the zeros of rational harmonic 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; Gravitational lensing; Critical curve and caustic; Cusp and fold points; Singular zeros

AMS Subject Classification (2010): 30D05, 31A05, 85A04

1 Introduction

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

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

(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) 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

* 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), namely

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

(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 $\eta$ 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 $\eta$. Using explicit computations, Schneider and Weiss studied this behavior for two point-masses, i.e., $n = 2$ in [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 $\eta$ on the zeros of rational harmonic functions of the form

$$
  f_\eta(z) = f(z) - \eta, \quad \text{where} \quad f \text{ is as in (1)}.
$$

In particular, we study the behavior of the zeros when $\eta$ 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 $f_\eta$ by $4\deg(r) - 6$. Thus, the effect of varying $\eta$ 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 $\eta$ 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 $f(z) = r(z) - \bar{z}$ with $\deg(r) \geq 2$ be given. Using the Wirtinger derivatives $\partial_z$ and $\partial_z$ we can write the Jacobian of $f$ as

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

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 $C$. 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 $\Omega_+ \cup \Omega_-$, respectively, so that we have the disjoint partitioning $\mathbb{C} = C \cup \Omega_+ \cup \Omega_-$. Each closed curve in the set $C$ is called a critical curve of $f$.

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

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

and hence the condition $r''(z) \neq 0$ for all $z \in C$ 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 $\partial A_j \subseteq C$, and either $A_j \subseteq \Omega_+$ or $A_j \subseteq \Omega_-$, for $j = 1, \ldots, m$, and we write

$$\mathcal{A} := \{A_1, \ldots, A_m\}. \quad (3)$$

Exactly one of the sets $A_j$ is unbounded, and we sometimes denote this set by $A_\infty$. 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 $1/r'$. 

3
The elements of the set $f(C)$ are called the **caustic points** of $f$, and for each critical curve $\Gamma$, 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 $z_0 \in C$ is given by

$$g(t) = ht + z_0,$$

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 $z_0 \in C$ 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

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

which shows that the tangent direction at the caustic point $f(z_0) \in f(C)$ 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

$$\text{Im}(e^{i\varphi/2}h) = \text{Im}(ie^{3i\varphi/2}r'(z_0)) = 0,$$
or, equivalently,
\[
\text{Re} \left( \frac{r''(z_0)}{r'(z_0)^{3/2}} \right) = 0 \iff \text{Re} \left( \frac{r'(z_0)^{3/2}}{r''(z_0)} \right) = 0,
\]
where the equivalence is defined since we assume that \( f \) is non-degenerate.

Let us summarize these considerations.

**Lemma 2.1.** Let \( z_0 \in \mathbb{C} \) be a critical point of \( f(z) = r(z) - z \). Then the caustic point \( f(z_0) \in f(\mathbb{C}) \) is a cusp point if and only if (5) holds. (Each other caustic point \( f(z_0) \in f(\mathbb{C}) \) is called a fold point.)

Petters and Witt \[16\] showed that if \( r \) is as in (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 \( f(z_0) = r(z_0) - z_0 = 0 \). We call \( z_0 \) a sense-preserving, sense-reversing, or singular zero of \( f \), if \( z_0 \) is an element of \( \Omega_+ \), \( \Omega_- \), or \( \mathbb{C} \), 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 \( \varepsilon \). 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 \( f(z) = r(z) - z \) 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 \( \eta \) is a caustic point of \( f \).

**Proof.** If \( z_0 \in \mathbb{C} \) is a singular zero of \( f_\eta \), then \( z_0 \in \mathbb{C} \) and \( f(z_0) - \eta = 0 \), or \( f(z_0) = \eta \), which means that \( \eta \) 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 \( z_0 \in \mathbb{C} \), which means that \( z_0 \) is a singular zero of \( f_\eta \).

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., \( \eta \notin f(\mathbb{C}) \), 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 \( \Gamma \) be a closed Jordan curve, and let \( f \) be a function that is continuous and nonzero on \( \Gamma \). Then the winding of \( f \) on \( \Gamma \) is defined as the change in
the argument of $f(z)$ as $z$ travels once around $\Gamma$ in the positive direction, divided by $2\pi$, 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 $\text{ind}(f; z_0) := V(f; \Gamma)$, where $\Gamma$ 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 $\Gamma$, as long as $z_0$ is the only exceptional point of $f$ in $\text{int}(\Gamma)$, the interior of $\Gamma$. If there are several (isolated) exceptional points in $\text{int}(\Gamma)$, we have the following theorem.

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

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

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 $\Gamma$ and has no singular zero in $\text{int}(\Gamma)$, then

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

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

6
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 \( \Gamma \) 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

\[
f(z) = r(z) - \tau \quad \text{with} \quad \deg(r) \geq 2 \quad \text{and} \quad \lim_{|z| \to \infty} |f(z)| = \infty. \tag{6}
\]

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

(i) \( N(f_\eta; B_\varepsilon(v_j)) = N +(f_\eta; B_\varepsilon(v_j)) = \mu_j \) for \( j = 1, \ldots, m \),

(ii) \( z_j \in A_\infty \setminus \bigcup_{k=1}^n B_\varepsilon(v_k) \) for \( j = \deg(q) + 1, \ldots, \deg(q) + k \).

**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 \( r(z_j) - \tau_j \gg 1 \), which can happen when the zero \( z_j \) of \( f_\eta \) 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 \( j = 1, \ldots, m \),

(b) \( B_\varepsilon(z_j) \cap B_\varepsilon(z_\ell) = \emptyset \) for all \( j, \ell \in \{1, \ldots, m\} \) 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 \( \mu_j \) 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 \( B_{\delta}(-\eta) \cap B_{\varepsilon}(v_j) = \emptyset \) for all \( j = 1, \ldots, m \). Using the function \( g(z) := -\varepsilon - \eta \), which has \(-\eta\) as its only zero, we obtain

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

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 \( f_\eta \), which is contained in the set \( A_\infty \).

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

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

\[
r(z) = cz^k + \rho(z) + \tilde{r}(z), \quad k = \deg(p) - \deg(q),
\]

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

(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 (7) with \( |c| \neq 1 \). We will first show that for each \( \eta \) the function \( g(z) := cz - \eta - \varepsilon \) has exactly one zero. Writing \( c = \text{Re}(c) + i\text{Im}(c) \) and \( z = x + iy \), the equation \( g(z) = 0 \) can be written as

\[
\begin{bmatrix}
\text{Re}(c) - 1 & -\text{Im}(c) \\
\text{Im}(c) & \text{Re}(c) + 1
\end{bmatrix}
\begin{bmatrix}
x \\ y
\end{bmatrix} = \frac{\eta}{|c|}.
\]
The determinant of the matrix is $\text{Re}(c)^2 - 1 + \text{Im}(c)^2 = |c|^2 - 1 \neq 0$. Denoting the unique zero of $g$ by $z_g$ and using that $\lim_{|z| \to \infty} |\tilde{R}(z)| = 0$, we can choose $\eta > 0$ sufficiently large so that $|\tilde{R}(z)| < |cz - \eta - \overline{z}|$ 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 $B_\delta(z_g) \cap B_\varepsilon(v_j) = \emptyset$ for all $j = 1, \ldots, m$. We then get

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

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 (7) we can choose $\eta > 0$ sufficiently large and $\delta > 0$ so that $B_\delta(\eta_j) \subset A_\infty$ for all $j = 1, \ldots, k$, and $B_\delta(\eta_j) \cap B_\varepsilon(v_j) = \emptyset$ for all $j \neq \ell$, as well as $|\rho(z) + \tilde{R}(z) - \overline{z}| < |g(z)|$ for all $z \in \partial B_\delta(\eta_j)$, $j = 1, \ldots, k$. Therefore

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

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 (6), 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

$$\left(\bigcup_{j=1}^M B_\varepsilon(z_j)\right) \cap \mathbb{C} = \emptyset, \quad \text{and} \quad B_\varepsilon(z_j) \cap B_\varepsilon(z_k) = \emptyset \quad \text{for all} \quad j, k \in \{1, \ldots, M\} \quad \text{with} \quad j \neq k.$$

If $\eta \in \mathbb{C}$ satisfies

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

then the following properties hold:

(i) For each $j = 1, \ldots, m$ the functions $f$ and $f_\eta$ 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 $f_\eta$ have the same orientation on each set $B_\varepsilon(z_j)$. Moreover, for all $z \in \partial B_\varepsilon(z_j)$ we have

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

(8)
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 \( f_\eta \) coincide, the assertion follows from Corollary 2.5.

(ii) We know from (i) that \( f_\eta \) has exactly one zero in each of the sets \( B_{\varepsilon}(z_j) \), \( j = 1, \ldots, m \). If \( f_\eta \) has an additional zero \( z_0 \in \mathbb{C} \setminus \bigcup_{j=m+1}^{M} B_{\varepsilon}(z_j) \), then we can choose an \( \varepsilon > 0 \) such that \( f \) is nonzero on \( B_{\varepsilon}(z_0) \). Then (8) holds for all \( z \in \partial B_{\varepsilon}(z_0) \), and Theorem 2.6 yields \( V(f; \partial B_{\varepsilon}(z_0)) = V(f_\eta; \partial B_{\varepsilon}(z_0)) \), which is a contradiction, since \( f \) and \( f_\eta \) have the same number of poles, but \( f \) has no zeros in \( B_{\varepsilon}(z_0) \).

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 \( f_\eta \) remains constant as long as the shift \( \eta \) does not cross a caustic of \( f \). Our proof is based on the following two lemmas.

**Lemma 3.3.** If \( \Gamma \) 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 \( 0 \leq \lambda \leq 1 \), 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 \( 0 \leq \lambda \leq 1 \), and Proposition 2.2 implies that \( f_{\lambda \eta_1}(z) = f(z) - \lambda \eta_1 \neq 0 \) holds for all \( 0 \leq \lambda \leq 1 \) and \( z \in \Gamma \).

By construction and the triangle inequality we have

\[
\eta_1 = |f(z) - f_{\eta_1}(z)| \leq |f(z)| + |f_{\eta_1}(z)| \quad \text{for all } z \in \mathbb{C}. \tag{9}
\]

If equality holds in (9) 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 \( |f(z_0)| < \mu \eta_1 \), 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 (9), and hence \( V(f_{\eta_1}; \Gamma) = V(f; \Gamma) \) by Theorem 2.6. \( \square \)

**Lemma 3.4.** If \( \eta_1, \eta_2 \in \mathbb{C} \) are such that \( \lambda \eta_1 + (1-\lambda)\eta_2 \notin f(\mathbb{C}) \) holds for all \( 0 \leq \lambda \leq 1 \), 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 \( f_{\eta_1} \) and \( f_{\eta_2} \) 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 ⊂ int(Γ∗). If f is sense-preserving on A, then Corollary 2.5 and Lemma 3.3 imply

\[ N(f_\eta_1; A) = V(f_\eta_1; \Gamma^*) - \sum_{\Gamma \subset \partial A \text{ crit. curve}, \ \text{int}(\Gamma) \cap A = \emptyset} V(f_\eta_1; \Gamma) - P(f; \text{int}(\Gamma)) = N(f_\eta_2; A). \]

If f is sense-reversing on A, then

\[ N(f_\eta_1; A) = \sum_{\Gamma \subset \partial A \text{ crit. curve}, \ \text{int}(\Gamma) \cap A = \emptyset} V(f_\eta_1; \Gamma) - P(f; \text{int}(\Gamma)) = N(f_\eta_2; A). \]

Using an additional artificial curve \( \partial B_R(0) \) for a sufficiently large \( R > 0 \), containing all zeros and poles of \( f_\eta_1 \) and \( f_\eta_2 \) in its interior, we obtain the equality for the set \( A_\infty \). Finally, \( N(f_\eta_1) = N(f_\eta_2) \) holds since \( f_\eta_1 \) and \( f_\eta_2 \) 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 \( \psi([0, 1]) \cap f(\mathbb{C}) = \emptyset \). Since \( \mathbb{C} \setminus f(\mathbb{C}) \) is open, we can approximate \( \psi([0, 1]) \) arbitrarily closely by a polygonal chain in \( \mathbb{C} \setminus f(\mathbb{C}) \); 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 2: Illustration of Theorem 3.5](image)
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 $\eta_1$ and $\eta_2$. Since there is a continuous path from $\eta_1$ to $\eta_2$, which does not cross a caustic of $f$ (see the plot on the right), $f_{\eta_1}$ and $f_{\eta_2}$ 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 $\tilde{\eta} \in \mathbb{C}$.

Figure 3: Left and middle: Critical curves (solid), sense-preserving zeros (▲), sense-reversing zeros (▼), and poles (■) for $f_{\eta_1}$ (left) and $f_{\eta_2}$ (middle). Right: Caustics and the corresponding shift (dashed).

Remark 3.6. A rational harmonic function $f$ as in (6) is called extremal, when it has the maximum number of $5 \deg(r) - 5$ zeros. As mentioned in the Introduction, an explicit construction of Rhie [17] yields an extremal function $f$ with $r$ as in (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 $f_{\eta}$ has exactly $n + 1$ zeros, namely $n$ zeros close to the poles of $f$, and one zero in $A_{\infty}$. Thus, $f_{\eta}$ 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 $f_{\eta}$ 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 (6). Let $z_0 \in \mathbb{C}$ 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
\[ r(z_0) - \eta = z_0, \] we then have
\[
 f_\eta(z) = r'(z_0)(z - z_0) + \frac{r''(z_0)}{2}(z - z_0)^2 - (\overline{z} - z_0) + \sum_{k=3}^{\infty} \frac{r^{(k)}(z_0)}{k!}(z - z_0)^k.
\] (10)

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 \( f_\eta \) that we will derive in the following. Under our assumption we can write (10) as
\[
 f_\eta(z) = T(z) + R(z),
\] where
\[
 T(z) := dz^2 + z - \overline{z}, \quad d := \frac{r''(0)}{2}, \quad R(z) := \sum_{k=3}^{\infty} \frac{r^{(k)}(0)}{k!}z^k. \] (12)

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 \( f_\eta \) 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
\[
 z = \begin{cases} 
 0, & \text{if } \delta = 0, \\
 \pm \sqrt{\delta}, & \text{if } \delta > 0,
\end{cases}
\] (13)
and all non-real zeros of \( T_{\delta d} \) are given by
\[
 z = \begin{cases} 
 -d^{-1} \pm i\sqrt{|d^{-2}| - \delta}, & \text{if } |d^{-2}| \geq \delta \text{ and } \text{Re}(d^{-1})^2 \neq \delta, \\
 -d^{-1} - i\text{Im}(d^{-1}), & \text{if } |d^{-2}| \geq \delta \text{ and } \text{Re}(d^{-1})^2 = \delta, \text{ and } \text{Im}(d^{-1}) \neq 0.
\end{cases}
\] (14)

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

**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
\[
 0 = y^2 + 2\beta y - x^2 + \delta, \quad (15)
\]
\[
 0 = xy + \alpha y, \quad (16)
\]
which we need to solve for real $x$ and $y$.

If $y = 0$, then \([15]\) implies $x^2 = \delta$, where both $x$ and $\delta$ are real. Thus, all solutions of $T_{\delta d}(z) = 0$ with $\text{Im}(z) = 0$ 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 \([16]\) implies $x = -\alpha$, and substituting $x^2 = \alpha^2$ in \([15]\) yields

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

A solution of $T_{\delta d}(z) = 0$ with $\text{Im}(z) \neq 0$ 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 $\alpha^2 - \delta = \text{Re}(d^{-1})^2 - \delta \neq 0$, 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

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

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

**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 $\delta \neq \text{Re}(d^{-1})^2$, 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 $|z| \geq |\text{Re}(d^{-1})| > 0$, independently of the size of $\delta$. Thus, if $\text{Re}(d) \neq 0$, 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 \([6]\) 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 $\tilde{\eta} \in \mathbb{C}$, such that for all $0 < \alpha \leq 1$ we have

(i) $N(f_{\eta + \alpha \tilde{\eta}}) = N(f_{\eta}) + 1 = N(f_{\eta - \alpha \tilde{\eta}}) + 2$,

(ii) $N(f_{\eta + \alpha \tilde{\eta}}; A) = N(f_{\eta}; A) = N(f_{\eta - \alpha \tilde{\eta}}; A)$ for all $A \in \mathcal{A} \setminus \{A_+, A_-\}$,

(iii) $N(f_{\eta + \alpha \tilde{\eta}}; A_+) = N(f_{\eta - \alpha \tilde{\eta}}; A_+) + 1$,

(iv) $N(f_{\eta + \alpha \tilde{\eta}}; A_-) = N(f_{\eta - \alpha \tilde{\eta}}; A_-) + 1$,

(v) $N_s(f_{\eta + \alpha \tilde{\eta}}) = N_s(f_{\eta - \alpha \tilde{\eta}}) = 0$, and $N_s(f_{\eta}) = 1$.

**Proof.** We will write $f_\eta$ as in \([11]\)--\([12]\), 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 $\text{Re}(d) \neq 0$; see Lemma \([2.1]\).
We know that if \( \tilde{\delta} > 0 \) is small enough, then there exists an \( \tilde{\varepsilon} \), depending on \( \tilde{\delta} \) and with \( \tilde{\varepsilon} > \sqrt{\tilde{\delta}} > 0 \), such that the only zeros of \( T_{+\tilde{\delta}d} \) in the open disk \( B_{\tilde{\varepsilon}}(0) \) are the two real zeros \( z = \pm \sqrt{\tilde{\delta}} \). The function \( T_{-\tilde{\delta}d} \) has no zeros in that disk. Moreover, by shrinking \( \tilde{\varepsilon} \) and \( \tilde{\delta} \) if necessary, we can assume that \( f_{\eta \pm \delta d} \) has no pole in \( B_{\tilde{\varepsilon}}(0) \), since \( |f_{\eta \pm \delta d}(0)| = |\delta d| < \infty \).

The orientation of \( T_{+\tilde{\delta}d} \) is determined by its Jacobian

\[
J_{T_{+\tilde{\delta}d}}(z) = |2dz + 1|^2 - 1.
\]

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

The main idea now is to suitably choose \( \varepsilon \) and \( \delta \) with \( \varepsilon > \sqrt{\delta} > 0 \), by possibly further shrinking the values \( \tilde{\varepsilon} \) and \( \tilde{\delta} \) 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\( \Gamma^+ := (B^+, B_{\varepsilon}(0) \cap \mathcal{C}) \) and \( \Gamma^- := (B^-, B_{\varepsilon}(0) \cap \mathcal{C}) \), (17)

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

\[
|f_{\eta \pm \delta d}(z) - T_{\pm \delta d}(z)| = |R(z)| < |f_{\eta \pm \delta d}(z)| + |T_{\pm \delta d}(z)|
\]

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 \( \varepsilon \) 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

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

We can assume that \( 0 < \varepsilon^* \leq \tilde{\varepsilon} \), 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 \( B_0(0) = \emptyset \). Thus, for every continuous and strictly monotonically increasing function\( \psi : [0, \infty) \to [0, \infty) \) with \( \psi(0) = 0 \),
there exists a $\delta^* > 0$, such that $\varepsilon(\delta^*) > \psi(\delta^*)$. Using the function $\psi(t) := \sqrt{t}$ yields parameters $\delta^*$ and $\varepsilon(\delta^*)$ with $\varepsilon > \varepsilon(\delta^*) > \sqrt{\delta^*} > 0$, so that only the two real zeros $z_\pm = \pm \sqrt{\delta^*}$ of $T_{t^*}d$ lie in the disk $B_{\varepsilon(\delta^*)}(0)$, and $T_{t^*}d$ is differently oriented at these zeros.

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

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

We also have to verify inequality (18) on $B_{\varepsilon(\delta^*)}(0) \cap \mathcal{C}$. Using (1) with our assumptions $z_0 = 0$, $r'(0) = 1$, and $d = r''(0)/2$, we see that this curve is given by

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

A straightforward computation shows that

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

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

$$|T_{t^*}d(ht)| = \left| -\bar{d}t^2 + 2i \frac{\text{Re}(d)}{|d|} t + \delta^*d \right| \geq \left| \text{Re}(d) \right| \cdot |t^2 + \delta^*| \geq \frac{\left| \text{Re}(d) \right| |\delta^*|}{2} > c_1|t^3| \geq |R(ht)|,$$

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

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,

$$|T_{t^*}d(ht)| = \left| -\bar{d}t^2 + 2i \frac{\text{Re}(d)}{|d|} t + \delta^*d \right| \geq \left| \text{Im}(d) (t^2 + \delta^*) + 2 \frac{\text{Re}(d)}{|d|} t \right| \geq c_2 \sqrt{\delta^*} > c_3|t^3| \geq |R(ht)|,$$

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

Together we have

$$|T_{t^*}d(ht)| > |R(ht)| \quad \text{for} \quad |t| < \varepsilon(\delta^*). \quad (19)$$  

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 (19) for a possibly smaller $\varepsilon^* > 0$. Hence,

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

16
holds for all $t$ with $|g(t)| \leq \varepsilon(\delta^*)$. Consequently, (18) is fulfilled for all $z \in B_{\varepsilon(\delta^*)}(0) \cap C$.

In summary, we can apply Theorem 2.6 on $\Gamma^+$ and $\Gamma^-$, see (17). With Corollary 2.5 this yields

\[
N(f_{\eta + \delta^*d}; \text{int}(\Gamma^+)) = V(f_{\eta + \delta^*d}; \Gamma^+) = V(T_{+\delta^*d}; \Gamma^+) = 1, \\
N(f_{\eta + \delta^*d}; \text{int}(\Gamma^-)) = -V(f_{\eta + \delta^*d}; \Gamma^-) = -V(T_{+\delta^*d}; \Gamma^-) = 1, \\
N(f_{\eta - \delta^*d}; \text{int}(\Gamma^+)) = V(f_{\eta - \delta^*d}; \Gamma^+) = V(T_{-\delta^*d}; \Gamma^+) = 0, \\
N(f_{\eta - \delta^*d}; \text{int}(\Gamma^-)) = -V(f_{\eta - \delta^*d}; \Gamma^-) = -V(T_{-\delta^*d}; \Gamma^-) = 0.
\]

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 $\tilde{\eta} := \delta^*d$. 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 $f$. 

![Figure 4: Local behavior near fold points.](image)

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 $z_0 \in C$ 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 $\eta$, 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
η − δ∗d to η + δ∗d. Using Theorem 3.5, we easily see that crossing the caustic 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 $f_\eta$ has no zero close to the critical point $z_0$. For $\eta = f(z_0)$ there is exactly one (singular) zero of $f_\eta$, and after $\eta$ crossed the caustic of $f$, a pair of differently oriented zeros of $f_\eta$ 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 $f_{\eta_1}$ and $f_{\eta_2}$. On the right we plot the caustics and one possible path from $\eta_1$ to $\eta_2$. On every path from $\eta_1$ to $\eta_2$ 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 $\eta_1$ to $\eta_2$, 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 $-d$ 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.

![Figure 5: Left and middle: Critical curves (solid), sense-preserving zeros (▲), sense-reversing zeros (▼), and poles (■) for $f_{\eta_1}$ (left) and $f_{\eta_2}$ (middle). Right: Caustics and the corresponding shift (dashed).](image)

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 $g(z) - \bar{z}$ is univalent in $D$. 

18
If $g$ is analytic and $|g'(z)| < 1$ in a star domain $D$ with base point $z_0,$ then we can apply this proposition on the lines from $z_0$ to any point of $D,$ which implies that $g(z) - z$ attends the value $g(z_0) - z_0$ exactly once in $D.$ This fact will be used in the proof of the next theorem.

**Theorem 4.5.** Let $f$ be as in (6) 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 $\tilde{\eta} \in \mathbb{C}$ and $b_+, b_- \in \{0, 1\}$ with $b_+ + b_- = 1,$ such that for all $0 < \alpha \leq 1$ we have

(i) $N(f_{\eta + \alpha \tilde{\eta}}; A_+) = N(f_{\eta} - \alpha \tilde{\eta}; A_+) + 1,$

(ii) $N(f_{\eta + \alpha \tilde{\eta}}; A_-) = N(f_{\eta} - \alpha \tilde{\eta}; A_-) + 1,$

(iii) $N_s(f_{\eta + \alpha \tilde{\eta}}) = N_s(f_{\eta} - \alpha \tilde{\eta}) = 0,$ and $N_s(f_{\eta}) = 1.$

**Proof.** The equalities

\[ N(f_{\eta + \alpha \tilde{\eta}}; A_+) = N(f_{\eta} - \alpha \tilde{\eta}; A_+) + 1, \]

\[ N(f_{\eta + \alpha \tilde{\eta}}; A_-) = N(f_{\eta} - \alpha \tilde{\eta}; A_-) + 1, \]

\[ N_s(f_{\eta + \alpha \tilde{\eta}}) = N_s(f_{\eta} - \alpha \tilde{\eta}) = 0, \]

already follow from Theorems 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 $\text{Re}(d) = 0$ (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

\[ z = -d^{-1} \pm i \sqrt{|d^{-2}| - \delta}, \tag{20} \]
and \( T_{-\delta d} \) has the two purely imaginary zeros
\[
z = -d^{-1} \pm i\sqrt{|d^{-2}| + \delta}. \tag{21}
\]
Only one of the two zeros in (20) and in (21) is sufficiently close to \( z = 0 \), and the sign of \( \text{Im}(d^{-1}) \) determines which one it is: If \( \text{Im}(d^{-1}) > 0 \), then the zero of interest of \( T_{\pm\delta d} \) is
\[
z_\pm := -d^{-1} + i\sqrt{|d^{-2}| + \delta}
\]
since then \( |z_\pm| < \sqrt{\delta} \), while the other zero satisfies
\[
| -d^{-1} - i\sqrt{|d^{-2}| + \delta} | > |d^{-1}| > \sqrt{\delta}.
\]
From
\[
|\partial_\xi T_{\pm\delta d}(z_\pm)| = |2dz_\pm + 1| = |2d(-d^{-1} + i\sqrt{|d^{-2}| + \delta}) + 1| = | -1 + 2d\sqrt{|d^{-2}| + \delta} | = | -1 + 2\sqrt{1 + \delta|d|^{-2}} |
\]
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 \( \text{Im}(d^{-1}) < 0 \) we get an analogous result, but then the zeros in (20) and (21) change their roles, i.e., \( z = -d^{-1} - i\sqrt{|d^{-2}| + \delta} \) is close to zero, and \( z = -d^{-1} + i\sqrt{|d^{-2}| + \delta} \) 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 \( \partial B_\varepsilon(z_\pm) \) for an appropriately chosen \( \tilde{\varepsilon} > 0 \). For each \( \varphi \in [0, 2\pi) \) we have
\[
T_{\pm\delta d}(z_\pm + \tilde{\varepsilon}e^{i\varphi}) = \tilde{\varepsilon} (2dz_\pm e^{i\varphi} + d\tilde{\varepsilon}e^{2i\varphi} + 2i \text{Im}(e^{i\varphi})).
\]
For a sufficiently small \( \delta > 0 \), which determines \( z_\pm \), we now set \( \tilde{\varepsilon} := |z_\pm| \), and we assume that \( 2|d| \tilde{\varepsilon} \leq 1 \). Then
\[
|T_{\pm\delta d}(z_\pm + \tilde{\varepsilon}e^{i\varphi})| \geq \tilde{\varepsilon} (2|d| \tilde{\varepsilon} |\text{Re}(e^{i\varphi})| - |d| \tilde{\varepsilon}) 
\geq \tilde{\varepsilon} (2|d| \tilde{\varepsilon} |\text{Re}(e^{i\varphi})| + |d| \tilde{\varepsilon} |\text{Im}(e^{i\varphi})| - |d| \tilde{\varepsilon}) 
\geq \tilde{\varepsilon} (2|d| \tilde{\varepsilon} - |d| \tilde{\varepsilon}) + |d| \tilde{\varepsilon}^2 
\geq c|z_{\text{Im}} + \tilde{\varepsilon}e^{i\varphi}|^3 \geq |R(z_{\text{Im}} + \tilde{\varepsilon}e^{i\varphi})|,
\]
for some constant \( c > 0 \). Thus, we have
\[
|f_{\eta\pm\delta d}(z) - T_{\pm\delta d}(z)| = |R(z)| < |T_{\pm\delta d}(z)| + |f_{\eta\pm\delta d}(z)| \quad \text{for all } z \in \partial B_{\varepsilon}(z_\pm).
\]
Using Theorem 2.6 gives
\[
V(f_{\eta\pm\delta d}; \partial B_{\varepsilon}(z_\pm)) = V(T_{\pm\delta d}; \partial B_{\varepsilon}(z_\pm)),
\]
and, as a consequence,
\[ N(f_{\eta \pm \delta d}; B_{\tilde{e}}(z_{\pm})) = N(T_{\pm \delta d}; B_{\tilde{e}}(z_{\pm})) = 1. \]

In the following we denote by \( \tilde{z}_{\pm} \) the zero of \( f_{\eta \pm \delta d} \) corresponding to the zero \( z_{\pm} \) of \( T_{\pm \delta d} \). By construction, \( \tilde{z}_{+} \) is a sense-reversing zero of \( f_{\eta + \delta d} \) and \( \tilde{z}_{-} \) 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 \( \partial B_{e}(0) \) and the zeros \( \tilde{z}_{\pm} \) 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 \( f_{\eta} \) has no poles in \( B_{\varepsilon}(0) \). Furthermore we define
\[
\tilde{\delta} := \min \left\{ \frac{|f_{\eta}(z)|}{|d|} : z \in \partial B_{\varepsilon}(0) \right\} \quad \text{and} \quad \delta := \frac{1}{2} \min \{\tilde{\delta}, \varepsilon^{2}\}.
\]

Hence, we have
\[
|f_{\eta}(z) - f_{\eta \pm \delta d}(z)| = \delta|d| < |f_{\eta}(z)| + |f_{\eta \pm \delta d}(z)| \quad \text{for all} \quad z \in \partial B_{\varepsilon}(0).
\]

With Theorem 2.6 we get
\[ V(f_{\eta \pm \delta d}; \partial B_{\varepsilon}(0)) = V(f_{\eta}; \partial B_{\varepsilon}(0)) \quad \text{and} \quad \tilde{z}_{\pm} \in B_{\varepsilon}(0). \quad (22) \]

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 \( \tilde{z}_{+} \), the function \( f_{\eta + \delta d} \) has no other zero than \( \tilde{z}_{+} \) 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 \( \eta \), has either no \( (b_{-} = 0) \) or two \( (b_{-} = 1) \) zeros in \( B_{\varepsilon}(0) \cap \Omega_{-} \): cf. Theorem 4.3. Furthermore, because of (22), 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 \( z_{0} \in \mathbb{C} \), 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 \( \eta \) is shifted along the dotted line. Coming from the right the function has only one sense-preserving zero close to \( z_{0} \). When \( \eta \) reaches the caustic, the unique zero becomes singular. When \( \eta \) 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 (6), and $z_0 \in \mathbb{C}$ is a singular zero of $f$, then

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

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 $B_\varepsilon(z_0)$. 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.

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 $\rho = \frac{3}{2}$, and plot the zeros of $f_\eta$ for several constant shifts $\eta$ in Figure 8.

We know from Theorem 3.1 that for $|\eta| \gg 1$ the function $f_\eta$ has 3 zeros close to its 3 poles, and one zero in the set $A_\infty$. This can be observed for the shift $\eta_1$. The shift from $\eta_1$ to $\eta_2$ 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 $\eta_2$ to $\eta_3$. Finally, the shift from $\eta_3$ to $\eta_4$ results in an additional pair of zeros at the inner critical curve. Note that $N(f_{\eta_4}) = 10 = 5n - 5$. Hence $f_{\eta_4}$ is an extremal rational harmonic function, and it has $6 = 4n - 6$ more zeros than $f_{\eta_1}$; 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 (6), 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 $f_{\eta}$ 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
If $f$ is singular, then for every $\varepsilon > 0$ there must exist an $\eta \in B_\varepsilon(0)$, such that $f_\eta$ is regular and $N(f_\eta) \leq N(f)$.

As another example we consider

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

and plot the results in Figure 9. For $\eta_1$ we again have 3 zeros close to the 3 poles and one zero in $A_\infty$, as shown by Theorem 3.1. The first caustic crossing from $\eta_1$ to $\eta_2$ results in one additional pair of zeros, but due to the curvature of the caustic, the shift from $\eta_2$ to $\eta_3$ reverses this effect. The last shift from $\eta_3$ to $\eta_4$ 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 $f(z) = h(z) - \bar{z}$ 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 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 \( r_1(z) - r_2(z) \) 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 (2015), p. 14.

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