The Journal of Geometric Analysis
Volume 17, Number 1, 2007

Univalence Criteria for Lifts of Harmonic Mappings to Minimal Surfaces

By *M. Chuaqui, P. Duren, and B. Osgood*

ABSTRACT. A general criterion in terms of the Schwarzian derivative is given for global univalence of the Weierstrass-Enneper lift of a planar harmonic mapping. Results on distortion and boundary regularity are also deduced. Examples are given to show that the criterion is sharp. The analysis depends on a generalized Schwarzian defined for conformal metrics and on a Schwarzian introduced by Ahlfors for curves. Convexity plays a central role.

1. Introduction

If a function f is analytic and locally univalent, its *Schwarzian derivative* is defined by

$$\mathcal{S}f = \big(f''/f'\big)' - \tfrac{1}{2}\big(f''/f'\big)^2 = f'''/f' - \tfrac{3}{2}\big(f''/f'\big)^2 .$$

The Schwarzian is invariant under postcomposition with Möbius transformations

$$T(z) = \frac{az+b}{cz+d}, \qquad ad - bc \neq 0;$$

that is, $\mathcal{S}(T \circ f) = \mathcal{S}(f)$. If g is any function analytic and locally univalent on the range of f, then

$$\mathcal{S}(g \circ f) = (\mathcal{S}(g) \circ f) f'^2 + \mathcal{S}(f). \tag{1.1}$$

In particular, $\mathcal{S}(g \circ T) = (\mathcal{S}(g) \circ T)T'^2$, since $\mathcal{S}(T) = 0$ for every Möbius transformation T. For an arbitrary analytic function ψ, the functions f with Schwarzian $\mathcal{S}f = 2\psi$ are those of the form $f = w_1/w_2$, where w_1 and w_2 are linearly independent solutions of the linear differential equation $w'' + \psi w = 0$. It follows that Möbius transformations are the only functions with $\mathcal{S}f = 0$. More generally, if $\mathcal{S}f = \mathcal{S}g$, then $f = T \circ g$ for some Möbius transformation T.

In a groundbreaking article, Nehari [17] showed that estimates on the Schwarzian provide criteria for global univalence. He made the key observation that if a function f is analytic and

Math Subject Classifications. Primary: 30C99; secondary: 31A05, 53A10.
Key Words and Phrases. Harmonic mapping, Schwarzian derivative, curvature, minimal surface.
Acknowledgements and Notes. The authors were supported in part by FONDECYT Grant # 1030589.

ISSN 1050-6926

locally univalent in a simply connected domain Ω, with Schwarzian $\mathcal{S}f = 2\psi$, then f is globally univalent if and only if no solution of the differential equation $w'' + \psi w = 0$, other than the zero solution, vanishes more than once in Ω. The univalence problem then reduces to a question about differential equations that can be analyzed by means of the Sturm comparison theorem. Specifically, Nehari proved that if f is analytic and locally univalent in the unit disk $\mathbb{D}$, and if

$$|\mathcal{S}f(z)| \le \frac{2}{\left(1-|z|^2\right)^2}, \qquad z \in \mathbb{D}, \tag{1.2}$$

then f is univalent in $\mathbb{D}$. He also showed that the inequality

$$|\mathcal{S}f(z)| \le \frac{\pi^2}{2} \tag{1.3}$$

implies univalence. Later he showed [18] (see also [19]) that f is univalent under the general hypothesis

$$|\mathcal{S}f(z)| \le 2p(|z|), \tag{1.4}$$

where $p(x)$ is a positive, continuous, even function with the properties that $(1-x^2)^2 p(x)$ is nonincreasing on the interval $[0,1)$ and no nontrivial solution u of the differential equation $u'' + pu = 0$ has more than one zero in $(-1,1)$. We will refer to such functions $p(x)$ as *Nehari functions*. Conditions (1.2) and (1.3) are special cases of (1.4). Essén and Keogh [13] solved some extremal problems for analytic functions satisfying Nehari's general condition. Osgood and Stowe [22] developed a common generalization of Nehari's criteria, and others, involving the curvature of a metric.

The main purpose of the present article is to derive a corresponding univalence criterion for harmonic mappings, or rather for their lifts to minimal surfaces. In previous work [2, 3, 4] we have defined the Schwarzian derivative of a harmonic mapping and have developed some of its properties. It is natural to identify a harmonic mapping with its Weierstrass-Enneper lift to a minimal surface, and it is this lift whose univalence is implied by bounds on the Schwarzian derivative. For the underlying harmonic mappings, univalent or not, our criterion is also shown to imply estimates on distortion and properties of boundary regularity that are better than those known or conjectured (see [24] or [12]) for the full normalized class of univalent harmonic mappings. In this respect our investigation can be viewed as a harmonic analogue of earlier work on analytic functions by Gehring and Pommerenke [14] and Chuaqui and Osgood [7, 8, 9].

A planar harmonic mapping is a complex-valued harmonic function $f(z)$, $z = x+iy$, defined on some domain $\Omega \subset \mathbb{C}$. If Ω is simply connected, the mapping has a canonical decomposition $f = h + \overline{g}$, where h and g are analytic in Ω and $g(z_0) = 0$ for some specified point $z_0 \in \Omega$. The mapping f is locally univalent if and only if its Jacobian $|h'|^2 - |g'|^2$ does not vanish. It is said to be orientation-preserving if $|h'(z)| > |g'(z)|$ in Ω, or equivalently if $h'(z) \neq 0$ and the dilatation $\omega = g'/h'$ has the property $|\omega(z)| < 1$ in Ω.

According to the Weierstrass-Enneper formulas, a harmonic mapping $f = h + \overline{g}$ with $|h'(z)| + |g'(z)| \neq 0$ lifts locally to map into a minimal surface, Σ, described by conformal parameters if and only if its dilatation $\omega = q^2$, the square of a meromorphic function q. The Cartesian coordinates (U, V, W) of the surface are then given by

$$U(z) = \operatorname{Re}\{f(z)\}, \quad V(z) = \operatorname{Im}\{f(z)\}, \quad W(z) = 2\operatorname{Im}\left\{\int_{z_0}^{z} h'(\zeta)q(\zeta)\,d\zeta\right\}.$$

We use the notation

$$\widetilde{f}(z) = \big(U(z), V(z), W(z)\big)$$

for the lifted mapping of Ω into Σ. The height of the surface can be expressed more symmetrically as

$$W(z) = 2\,\mathrm{Im}\left\{\int_{z_0}^{z} \sqrt{h'(\zeta)g'(\zeta)}\,d\zeta\right\},$$

since a requirement equivalent to $\omega = q^2$ is that $h'g'$ be the square of an analytic function. The first fundamental form of the surface is $ds^2 = e^{2\sigma}|dz|^2$, where the conformal factor is

$$e^{\sigma} = |h'| + |g'|.$$

The Gauss curvature of the surface at a point $\widetilde{f}(z)$ for which $h'(z) \neq 0$ is

$$K = -e^{-2\sigma}\,\Delta\sigma = -\frac{4|q'|^2}{|h'|^2\left(1+|q|^2\right)^4}, \tag{1.5}$$

where Δ is the Laplacian operator. Further information about harmonic mappings and their relation to minimal surfaces can be found in [12].

For a harmonic mapping $f = h + \overline{g}$ with $|h'(z)| + |g'(z)| \neq 0$, whose dilatation is the square of a meromorphic function, we have defined [2] the *Schwarzian derivative* by the formula

$$\mathcal{S}f = 2\left(\sigma_{zz} - \sigma_z^2\right), \tag{1.6}$$

where

$$\sigma_z = \frac{\partial\sigma}{\partial z} = \frac{1}{2}\left(\frac{\partial\sigma}{\partial x} - i\,\frac{\partial\sigma}{\partial y}\right), \qquad z = x + iy.$$

Some background for this definition, on conformal Schwarzians, is discussed in Section 4. With $h'(z) \neq 0$ and $g'/h' = q^2$, a calculation (cf. [2]) produces the expression

$$\mathcal{S}f = \mathcal{S}h + \frac{2\overline{q}}{1+|q|^2}\left(q'' - \frac{q'h''}{h'}\right) - 4\left(\frac{q'\overline{q}}{1+|q|^2}\right)^2.$$

As observed in [2], the formula remains valid if ω is not a perfect square, provided that neither h' nor g' has a simple zero.

It must be emphasized that we are not requiring our harmonic mappings to be locally univalent. In other words, the Jacobian need not be of constant sign in the domain Ω. The orientation of the mapping may reverse, corresponding to a folding in the associated minimal surface. It is also possible for the minimal surface to exhibit several sheets above a point in the (U, V)-plane. Thus, the lifted mapping $\widetilde{f}$ may be univalent even when the underlying mapping f is not.

The following theorem gives a criterion for the lift of a harmonic map to be univalent.

Theorem 1.1. *Let $f = h + \overline{g}$ be a harmonic mapping of the unit disk, with $e^{\sigma(z)} = |h'(z)| + |g'(z)| \neq 0$ and dilatation $g'/h' = q^2$ for some meromorphic function q. Let $\widetilde{f}$ denote the Weierstrass-Enneper lift of f into a minimal surface Σ with Gauss curvature $K = K(\widetilde{f}(z))$ at the point $\widetilde{f}(z)$. Suppose that*

$$|\mathcal{S}f(z)| + e^{2\sigma(z)}\left|K\left(\widetilde{f}(z)\right)\right| \leq 2p(|z|), \qquad z \in \mathbb{D}, \tag{1.7}$$

for some Nehari function p. Then $\widetilde{f}$ is univalent in $\mathbb{D}$.

When f is analytic and locally univalent in $\mathbb{D}$, the result reduces to Nehari's theorem cited earlier, since the minimal surface Σ is then a plane with $K = 0$.

Theorem 1.1 is sharp, but to show this, and to describe the extremal mappings, we need to know that $\widetilde{f}$ has a continuous extension to the closed disk. We state this now as a theorem, although in fact we will obtain stronger results on the smoothness of the boundary function.

Theorem 1.2. *Under the hypotheses of Theorem 1.1, the Weierstrass-Enneper lift $\widetilde{f}$ has an extension to $\overline{\mathbb{D}}$ that is continuous with respect to the spherical metric.*

According to Theorem 1.2, the condition (1.7) implies that the extended mapping $\widetilde{f}$ sends the unit circle to a continuous closed curve Γ lying on the surface Σ in $\mathbb{R}^3 \cup \{\infty\}$. In fact, Γ is a simple closed curve, or equivalently the mapping $\widetilde{f}$ is univalent in the closed unit disk, except in special circumstances which we now describe. A harmonic mapping f satisfying (1.7) is said to be *extremal* if Γ is *not* a simple closed curve; that is, if $\widetilde{f}(\zeta_1) = \widetilde{f}(\zeta_2)$ for some pair of distinct points ζ_1 and ζ_2 on the unit circle $\partial\mathbb{D}$. One calls $P = f(\zeta_1) = f(\zeta_2)$ a *cut point* of Γ. The following theorem describes a characteristic property of extremal mappings.

Theorem 1.3. *Under the hypotheses of Theorem 1.1, suppose the closed curve $\Gamma = \widetilde{f}(\partial\mathbb{D})$ is not simple and let P be a cut point. Then there exists a Euclidean circle or line C such that $C \setminus \{P\}$ is a line of curvature of $\widetilde{f}(\mathbb{D})$ on the surface Σ. Furthermore, equality holds in (1.7) along $\widetilde{f}^{-1}(C \setminus \{P\})$.*

Theorems 1.1 and 1.3 will be proved in Section 3, Theorem 1.2 in Section 6. In the last section of the article we will construct some examples of extremal mappings illustrating the phenomenon in Theorem 1.3 and showing that the criterion in Theorem 1.1 is sharp, in fact best possible in some particular cases. We will also show in that section that such extremal lifts must actually map the disk into either a catenoid or a plane; this will be a consequence of Theorem 1.3 and a purely differential geometric property of the catenoid.

We think it is striking that the theorems for analytic functions generalize in this manner, and one cannot help but wonder what other related aspects of classical geometric function theory have counterparts for harmonic mappings or their lifts. For example, corresponding to the sufficient condition (1.2) is a necessary condition for univalence: If f is analytic and univalent in $\mathbb{D}$ then it is subject to the sharp inequality

$$|\mathcal{S}f(z)| \leq \frac{6}{\left(1-|z|^2\right)^2}, \qquad z \in \mathbb{D}.$$

This was shown by Kraus [16] and was rediscovered by Nehari [17]. The Koebe function, $k(z) = z/(1-z)^2$, has $(1-z^2)^2\mathcal{S}k(z) = -6$. For univalent harmonic mappings whose dilatation is a square we have shown in [5] that $(1-|z|^2)^2|\mathcal{S}f(z)|$ is at least finite, and we have been able to obtain the numerical bound

$$\left(1-|z|^2\right)^2|\mathcal{S}f(z)| \leq 19,407\,.$$

This is far from sharp. The horizontal shear of the Koebe function with dilatation $\omega(z) = z^2$ has

$$\mathcal{S}f = -4\left(\frac{1}{1-z} + \frac{\bar{z}}{1+|z|^2}\right),$$

and $(1-|z|^2)^2|\mathcal{S}f(z)| \leq 16$. Pressing the analogy, 16 is thus a reasonable conjecture for the sharp constant in general. We do not know how the geometry of the minimal surface associated with the lift might enter into a result of this type.

Of additional interest in this article is that our analysis involves not only the classical Schwarzian and its conformal generalization, but also a version of the Schwarzian introduced by Ahlfors for curves in $\mathbb{R}^n$, to which we now turn.

2. Ahlfors' Schwarzian and univalence along curves

Ahlfors [1] introduced a notion of Schwarzian derivative for mappings of a real interval into $\mathbb{R}^n$ by formulating suitable analogues of the real and imaginary parts of $\mathcal{S}f$ for analytic functions f. A simple calculation shows that

$$\operatorname{Re}\{\mathcal{S}f\} = \frac{\operatorname{Re}\{f'''\overline{f'}\}}{|f'|^2} - 3\,\frac{\operatorname{Re}\{f''\overline{f'}\}^2}{|f'|^4} + \frac{3}{2}\frac{|f''|^2}{|f'|^2}\,.$$

For mappings $\varphi : (a, b) \to \mathbb{R}^n$ of class C^3 with $\varphi'(x) \neq 0$, Ahlfors defined the analogous expression

$$S_1\varphi = \frac{\langle\varphi''', \varphi'\rangle}{|\varphi'|^2} - 3\frac{\langle\varphi'', \varphi'\rangle^2}{|\varphi'|^4} + \frac{3}{2}\frac{|\varphi''|^2}{|\varphi'|^2}\,, \tag{2.1}$$

where $\langle\cdot, \cdot\rangle$ denotes the Euclidean inner product and now $|\mathbf{x}|^2 = \langle\mathbf{x}, \mathbf{x}\rangle$ for $\mathbf{x} \in \mathbb{R}^n$. Ahlfors also defined a second expression analogous to $\operatorname{Im}\{\mathcal{S}f\}$, but this is not relevant to the present discussion.

Ahlfors' Schwarzian is invariant under postcomposition with Möbius transformations; that is, under every composition of rotations, magnifications, translations, and inversions in $\mathbb{R}^n$. Only its invariance under inversion

$$\mathbf{x} \mapsto \frac{\mathbf{x}}{|\mathbf{x}|^2}\,, \qquad \mathbf{x} \in \mathbb{R}^n\,,$$

presents a difficulty; this can be checked by straightforward but tedious calculation. It should also be noted that S_1 transforms as expected under change of parameters. If $x = x(t)$ is a smooth function with $x'(t) \neq 0$, and $\psi(t) = \varphi(x(t))$, then

$$S_1\psi(t) = S_1\varphi(x(t))\,x'(t)^2 + \mathcal{S}x(t)\,.$$

With the notation $v = |\varphi'|$, Chuaqui and Gevirtz [6] used the Frenet-Serret formulas to show that

$$S_1\varphi = \big(v'/v\big)' - \tfrac{1}{2}\big(v'/v\big)^2 + \tfrac{1}{2}v^2\kappa^2 = \mathcal{S}(s) + \tfrac{1}{2}v^2\kappa^2\,, \tag{2.2}$$

where $s = s(x)$ is the arclength of the curve and κ is its curvature. Our proof of Theorem 1.1 will be based on the following result, also due to Chuaqui and Gevirtz in [6].

Theorem A. *Let $p(x)$ be a continuous function such that the differential equation $u''(x) + p(x)u(x) = 0$ admits no nontrivial solution $u(x)$ with more than one zero in $(-1, 1)$. Let $\varphi : (-1, 1) \to \mathbb{R}^n$ be a curve of class C^3 with tangent vector $\varphi'(x) \neq 0$. If $S_1\varphi(x) \leq 2p(x)$, then φ is univalent.*

If the function $p(x)$ of Theorem A is even, as will be the case for a Nehari function, then the solution u_0 of the differential equation $u'' + pu = 0$ with initial conditions $u_0(0) = 1$ and $u_0'(0) = 0$ is also even, and therefore $u_0(x) \neq 0$ on $(-1, 1)$, since otherwise it would have at least two zeros. Thus, the function

$$\Phi(x) = \int_0^x u_0(t)^{-2}\,dt\,, \qquad -1 < x < 1\,, \tag{2.3}$$

is well defined and has the properties $\Phi(0) = 0$, $\Phi'(0) = 1$, $\Phi''(0) = 0$, $\Phi(-x) = -\Phi(x)$. The standard method of reduction of order produces the independent solution $u = u_0\Phi$ to $u'' + pu = 0$, and so $\mathcal{S}\Phi = 2p$. Note also that $S_1\Phi = \mathcal{S}\Phi$, since Φ is real-valued. Thus, $S_1\Phi = 2p$.

The next theorem, again to be found in [6], asserts that the mapping $\Phi : (-1, 1) \to \mathbb{R} \subset \mathbb{R}^n$ is extremal for Theorem A if $\Phi(1) = \infty$, and that every extremal mapping φ is then a Möbius postcomposition of Φ.

Theorem B. *Let $p(x)$ be an even function with the properties assumed in Theorem A, and let Φ be defined as above. Let $\varphi : (-1, 1) \to \mathbb{R}^n$ satisfy $S_1\varphi(x) \le 2p(x)$ and have the normalization $\varphi(0) = 0$, $|\varphi'(0)| = 1$, and $\varphi''(0) = 0$. Then $|\varphi'(x)| \le \Phi'(|x|)$ for $x \in (-1, 1)$, and φ has an extension to the closed interval $[-1, 1]$ that is continuous with respect to the spherical metric. Furthermore, there are two possibilities, as follows.*

(i) *If $\Phi(1) < \infty$, then φ is univalent in $[-1, 1]$ and $\varphi([-1, 1])$ has finite length.*

(ii) *If $\Phi(1) = \infty$, then either φ is univalent in $[-1, 1]$ or $\varphi = R \circ \Phi$ for some rotation R of $\mathbb{R}^n$.*

Note that in case (ii) the mapping Φ sends both ends of the interval to the point at infinity and is therefore not univalent in $[-1, 1]$. The role of Φ as an extremal for the harmonic univalence criterion (1.7) will emerge in the following sections.

3. Univalence and extremal functions: Proofs of Theorems 1.1 and 1.3

The proofs of Theorems 1.1 and 1.3 will be based on Theorems A and B. The following lemma makes the connection.

Lemma 3.1. *Let f be a harmonic mapping of the unit disk with nonvanishing conformal factor e^σ and dilatation q^2 for some meromorphic function q. Let $\widetilde{f}$ be the lift of f to a minimal surface Σ with Gauss curvature K. Then*

$$\begin{aligned} S_1\widetilde{f}(x) &= \mathrm{Re}\{\mathcal{S}f(x)\} + \tfrac{1}{2}e^{2\sigma(x)}\big|K\big(\widetilde{f}(x)\big)\big| + \tfrac{1}{2}e^{2\sigma(x)}\kappa_e\big(\widetilde{f}(x)\big)^2 \\ &\le \mathrm{Re}\{\mathcal{S}f(x)\} + e^{2\sigma(x)}\big|K\big(\widetilde{f}(x)\big)\big|, \qquad -1 < x < 1, \end{aligned}$$

where $\kappa_e(\widetilde{f}(x))$ denotes the normal curvature of the curve $\widetilde{f} : (-1, 1) \to \Sigma$ at the point $\widetilde{f}(x)$. Equality occurs at a point x if and only if the curve is tangent to a line of curvature of Σ at the point $\widetilde{f}(x)$.

Proof. According to the formula (2.2), we have

$$S_1\widetilde{f} = \big(v'/v\big)' - \tfrac{1}{2}\big(v'/v\big)^2 + \tfrac{1}{2}v^2\kappa^2,$$

where $v = e^\sigma$ is the conformal factor of the surface Σ and κ is the curvature of the curve $\widetilde{f} : (-1, 1) \to \Sigma$. The tangential and normal projections of the curvature vector

$$\frac{d\mathbf{t}}{ds} = \frac{d}{ds}\left(\frac{d\widetilde{f}}{ds}\right)$$

are the geodesic or intrinsic curvature κ_i and the normal or extrinsic curvature κ_e, respectively. Thus, $\kappa^2 = \kappa_i^2 + \kappa_e^2$. In a previous article [3], we related the geodesic curvature of the lifted curve to the curvature of an underlying curve C in the parametric plane. In the present context C is the linear segment $(-1, 1)$, with curvature zero, and so our formula ([3], Equation (4)) reduces to $e^{\sigma}\kappa_i = -\sigma_y$, where $\sigma_y = \partial\sigma/\partial y$. Therefore, we find that

$$\begin{aligned} S_1\widetilde{f} &= \sigma_{xx} - \tfrac{1}{2}\sigma_x^2 + \tfrac{1}{2}e^{2\sigma}\left(\kappa_i^2 + \kappa_e^2\right) \\ &= \tfrac{1}{2}\left(\sigma_{xx} - \sigma_{yy} - \sigma_x^2 + \sigma_y^2\right) + \tfrac{1}{2}\left(\sigma_{xx} + \sigma_{yy}\right) + \tfrac{1}{2}e^{2\sigma}\kappa_e^2 \\ &= \mathrm{Re}\,\{\mathcal{S}f\} + \tfrac{1}{2}e^{2\sigma}|K| + \tfrac{1}{2}e^{2\sigma}\kappa_e^2\,, \end{aligned}$$

in view of the expression (1.5) for the Gauss curvature K. Here we have used the fact that $K \le 0$ for a minimal surface. In fact, $K = k_1k_2$, where k_1 and k_2 are the principal curvatures, the maximum and minimum of the normal curvature κ_e of the surface as the tangent direction varies. But a minimal surface has mean curvature zero; that is, $k_1 + k_2 = 0$. This implies that $K \le 0$ and $\kappa_e^2 \le |k_1k_2| = |K|$. Consequently,

$$S_1\widetilde{f} \le \mathrm{Re}\,\{\mathcal{S}f\} + e^{2\sigma}|K|\,,$$

with equality if and only if the curve is tangent to a line of curvature of the surface Σ. In other words, equality occurs precisely when the curve is aligned in a direction of maximum or minimum normal curvature. □

Proof of Theorem 1.1. By Lemma 3.1, the inequality (1.7) implies that the curve $\widetilde{f}$: $(-1, 1) \to \Sigma$ satisfies the hypothesis $S_1\widetilde{f}(x) \le 2p(x)$ of Theorem A. Thus, Theorem A tells us that $\widetilde{f}$ is univalent in the interval $(-1, 1)$.

In order to conclude that $\widetilde{f}$ is univalent in $\mathbb{D}$, we adapt a clever argument due to Nehari [18]. We want to show that $\widetilde{f}(z_1) \ne \widetilde{f}(z_2)$ for any given pair of distinct points $z_1, z_2 \in \mathbb{D}$. But if f satisfies (1.7), then so does every rotation $f(e^{i\theta}z)$. Consequently, we may assume that the hyperbolic geodesic passing through the points z_1 and z_2 intersects the imaginary axis orthogonally. Let $i\rho$ denote the point of intersection, and observe that the Möbius transformation

$$T(z) = \frac{i\rho - z}{1 + i\rho z}\,, \qquad z \in \mathbb{D}\,, \tag{3.1}$$

maps the disk onto itself and preserves the imaginary axis, so that it maps the given geodesic onto the real segment $(-1, 1)$. Moreover, T is an involution, so that $T = T^{-1}$ and T also maps the segment $(-1, 1)$ onto the geodesic through z_1 and z_2. In particular, $T(x_1) = z_1$ and $T(x_2) = z_2$ for some points x_1 and x_2 in the interval $(-1, 1)$. The composite function $F(z) = f(T(z))$ is a harmonic mapping of the disk whose lift $\widetilde{F}$ again maps $\mathbb{D}$ to the minimal surface Σ. We claim that

$$|\mathcal{S}F(x)| + e^{2\tau(x)}\left|K\left(\widetilde{F}(x)\right)\right| \le 2p(x)\,, \qquad -1 < x < 1\,, \tag{3.2}$$

where $e^{\tau} = |H'| + |G'|$ is the conformal factor associated with $F = H + \overline{G}$. Indeed, by the chain rule,

$$e^{\tau(x)} = e^{\sigma(T(x))}\left|T'(x)\right|,$$

whereas $\mathcal{S}F(x) = \mathcal{S}f(T(x))\,T'(x)^2$ and $\widetilde{F}(x) = \widetilde{f}(T(x))$. Therefore, by virtue of the hypothesis (1.7), the claim (3.2) will be established if we can show that

$$\left|T'(x)\right|^2 p(|T(x)|) \le p(x)\,, \qquad -1 < x < 1\,. \tag{3.3}$$

But

$$\frac{\left|T'(z)\right|}{1 - |T(z)|^2} = \frac{1}{1 - |z|^2}\,, \qquad z \in \mathbb{D}\,,$$

and so (3.3) reduces to the inequality

$$\left(1-|T(x)|^2\right)^2 p(|T(x)|) \le \left(1-x^2\right)^2 p(x)\,, \qquad -1 < x < 1\,, \tag{3.4}$$

which follows from the assumption that $\left(1-x^2\right)^2 p(x)$ is nonincreasing on [0, 1), since p is an even function and an easy calculation shows that $|T(x)| > |x|$. This proves our claim (3.2).

Finally, we return to the remark made at the beginning of the proof. In view of Lemma 3.1 and Theorem A, the inequality (3.2) implies that $\widetilde{F}$ is univalent in $(-1, 1)$. Therefore, $\widetilde{F}(x_1) \neq \widetilde{F}(x_2)$, which says that $\widetilde{f}(z_1) \neq \widetilde{f}(z_2)$. This proves Theorem 1.1. □

Proof of Theorem 1.3. We start with the assumption that a harmonic mapping f satisfies the hypotheses of Theorem 1.1 for some Nehari function $p(x)$, and its lift $\widetilde{f}$ has the property $\widetilde{f}(\zeta_1) = \widetilde{f}(\zeta_2)$ for some pair of distinct points ζ_1 and ζ_2 on the unit circle. The first step is to show that after suitable modification of f we may take ζ_1 and ζ_2 to be diametrically opposite points. More precisely, we will show that some Möbius transformation T of $\mathbb{D}$ onto itself produces a harmonic mapping $F = f \circ T$ with a lift $\widetilde{F}(z) = \widetilde{f}(T(z))$ for which the inequality (3.2) holds and $\widetilde{F}(1) = \widetilde{F}(-1)$. To make the analysis clear, we will distinguish two cases.

Suppose first that $(1-x^2)^2 p(x)$ is constant. Then equality holds in (3.4), hence in (3.3), for every Möbius automorphism of $\mathbb{D}$. Consequently, every mapping $F = f \circ T$ satisfies the inequality (3.2), and we will have $\widetilde{F}(1) = \widetilde{F}(-1)$ if we choose the automorphism T such that $T(1) = \zeta_1$ and $T(-1) = \zeta_2$.

If $(1-x^2)^2 p(x)$ is not constant, we claim that necessarily $\zeta_1 = -\zeta_2$. If not, then after suitable rotation we may assume that

$$\mathrm{Im}\,\{\zeta_1\} = \mathrm{Im}\,\{\zeta_2\} > 0 \qquad \text{and} \qquad \mathrm{Re}\,\{\zeta_2\} = -\mathrm{Re}\,\{\zeta_1\} > 0\,.$$

Now let $i\rho$ be the point where the imaginary axis meets the hyperbolic geodesic from ζ_1 to ζ_2, and observe that the Möbius transformation (3.1) carries the real segment $(-1, 1)$ onto this geodesic, with $T(1) = \zeta_1$ and $T(-1) = \zeta_2$. As in the proof of Theorem 1.1, we arrive at the inequality (3.4), which implies (3.3) and therefore (3.2), but now the inequality (3.2) cannot reduce to equality throughout the interval $(-1, 1)$, since $(1-x^2)^2 p(x)$ is not constant. In view of Lemma 3.1, we conclude that

$$S_1\widetilde{F}(x) \le 2p(x)\,, \qquad -1 < x < 1\,,$$

with strict inequality in some part of the interval. In particular, $S_1\widetilde{F} \not\equiv 2p$. However, this conclusion stands in contradiction to Theorem B. Indeed, after postcomposition with a suitable Möbius transformation M of $\mathbb{R}^3$, we obtain a mapping

$$\varphi = M \circ \widetilde{F} : (-1, 1) \to \mathbb{R}^3$$

with the required normalization $\varphi(0) = 0$, $|\varphi'(0)| = 1$, and $\varphi''(0) = 0$ (cf. [6]). But $\varphi(1) = \varphi(-1)$, so Theorem B tells us that $\varphi = R \circ \Phi$ for some rotation R. Hence, $(R^{-1}M) \circ \widetilde{F} = \Phi$, and so $S_1\widetilde{F} = S_1\Phi = 2p$. This contradicts our earlier conclusion and shows that $\zeta_1 = -\zeta_2$. Thus, some rotation T of the disk produces a harmonic mapping $F = f \circ T$ whose lift $\widetilde{F} = \widetilde{f} \circ T$ satisfies (3.2) and $\widetilde{F}(1) = \widetilde{F}(-1)$.

In all cases we find that some Möbius transformation T of the disk onto itself produces a harmonic mapping $F = f \circ T$ that satisfies the inequality (3.2), and whose lift $\widetilde{F} = \widetilde{f} \circ T$ has the property $\widetilde{F}(1) = \widetilde{F}(-1)$. Thus, $\widetilde{F}$ maps the interval $[-1, 1]$ to a closed curve on the surface Σ. As previously indicated, Theorem B then shows that $\widetilde{F} = V \circ \Phi$ for some Möbius transformation

V of $\mathbb{R}^3$. It follows that $S_1\widetilde{F} = S_1\Phi = 2p$, and also that $\widetilde{F}$ maps $[-1, 1]$ to a Euclidean circle or line, since Φ maps $[-1, 1]$ onto the extended real line and Möbius transformations preserve circles. On the other hand, since F satisfies (3.2), Lemma 3.1 shows that

$$S_1\widetilde{F}(x) \le \operatorname{Re}\{\mathcal{S}F(x)\} + e^{2\tau(x)}\big|K\big(\widetilde{F}(x)\big)\big| \le 2p(x)\,, \qquad -1 < x < 1\,.$$

But $S_1\widetilde{F}(x) = 2p(x)$, so equality holds throughout. According to Lemma 3.1, this says that the circle $\widetilde{F}([-1, 1])$ is everywhere tangent to a line of curvature, so it is in fact a line of curvature of Σ. $\square$

4. Conformal Schwarzian

Results on extensions to the boundary and estimates on distortion for harmonic mappings satisfying the univalence criterion (1.7) depend upon inequalities derived from convexity. These in turn call on a generalized Schwarzian that is computed with respect to a conformal metric and on a second order differential equation associated with the Schwarzian. It is this 'conformal Schwarzian' when specialized to the lift of a harmonic mapping that produces the definition (1.6); see also [2]. The definition and properties are suggested by the classical case, and have analogues there, but the generalization must be framed in the terminology of differential geometry; see, for example, [20] for a very accessible treatment of the operations we use here. This section provides a brief summary of the generalized Schwarzian, with all definitions given for dimension 2. We refer to [23] for the higher dimensional setting, and to [9] for applications of convexity in 2 dimensions, similar to what we will do here for harmonic mappings.

Let $\mathbf{g}$ be a Riemannian metric on the disk $\mathbb{D}$. We may assume that $\mathbf{g}$ is conformal to the Euclidean metric, $\mathbf{g}_0 = dx \otimes dx + dy \otimes dy = |dz|^2$. Let ψ be a smooth function on $\mathbb{D}$ and form the symmetric 2-tensor

$$\operatorname{Hess}_{\mathbf{g}}(\psi) - d\psi \otimes d\psi\,. \tag{4.1}$$

Here Hess denotes the Hessian operator. For example, if $\gamma(s)$ is an arclength parametrized geodesic for $\mathbf{g}$, then

$$\operatorname{Hess}_{\mathbf{g}}(\psi)\big(\gamma', \gamma'\big) = \frac{d^2}{ds^2}(\psi \circ \gamma)\,.$$

The Hessian depends on the metric, and since we will be changing metrics we indicate this dependence by the subscript $\mathbf{g}$.

With some imagination the tensor (4.1) begins to resemble a Schwarzian; among other occurrences in differential geometry, it arises (in 2 dimensions) if one differentiates the equation that relates the geodesic curvatures of a curve for two conformal metrics. Such a curvature formula is a classical interpretation of the Schwarzian derivative, see [23] and [3]. The trace of the tensor is the function

$$\frac{1}{2}\big(\Delta_{\mathbf{g}}\psi - ||\operatorname{grad}_{\mathbf{g}}\psi||^2_{\mathbf{g}}\big)\,,$$

where again we have indicated by a subscript that the Laplacian, gradient, and norm all depend on $\mathbf{g}$. It turns out to be most convenient to work with a traceless tensor when generalizing the Schwarzian, so we subtract off this function times the metric $\mathbf{g}$ and define the *Schwarzian tensor* to be the symmetric, traceless, 2-tensor

$$B_{\mathbf{g}}(\psi) = \operatorname{Hess}_{\mathbf{g}}(\psi) - d\psi \otimes d\psi - \frac{1}{2}\big(\Delta_{\mathbf{g}}\psi - ||\operatorname{grad}_{\mathbf{g}}\psi||^2\big)\mathbf{g}\,.$$

Working in standard Cartesian coordinates one can represent $B_{\mathbf{g}}(\psi)$ as a symmetric, traceless 2×2 matrix, say of the form

$$\begin{pmatrix} a & -b \\ -b & -a \end{pmatrix} .$$

Further identifying such a matrix with the complex number $a+bi$ then allows us to associate the tensor $B_{\mathbf{g}}(\psi)$ with $a+bi$.

At each point $z\in\mathbb{D}$, the expression $B_{\mathbf{g}}(\psi)(z)$ is a bilinear form on the tangent space at z, and so its norm is

$$||B_{\mathbf{g}}(\psi)(z)||_{\mathbf{g}} = \sup_{X,Y} B_{\mathbf{g}}(\psi)(z)(X,Y)\,,$$

where the supremum is over unit vectors in the metric $\mathbf{g}$. If we compute the tensor with respect to the Euclidean metric and make the identification with a complex number as above, then

$$||B_{\mathbf{g_0}}(\psi)(z)||_{\mathbf{g_0}} = |a+bi|\,.$$

Now, if f is analytic and locally univalent in $\mathbb{D}$, then it is a conformal mapping of $\mathbb{D}$ with the metric $\mathbf{g}$ into $\mathbb{C}$ with the Euclidean metric. The pullback $f^*\mathbf{g}_0$ is a metric on $\mathbb{D}$ conformal to $\mathbf{g}$, say $f^*\mathbf{g}_0 = e^{2\psi}\mathbf{g}$, and the (conformal) Schwarzian of f is now defined to be

$$\mathcal{S}_{\mathbf{g}} f = B_{\mathbf{g}}(\psi)\,.$$

If we take $\mathbf{g}$ to be the Euclidean metric then $\psi = \log|f'|$. Computing $B_{\mathbf{g}_0}(\log|f'|)$ and writing it in matrix form as above results in

$$B_{\mathbf{g}_0}\big(\log|f'|\big) = \begin{pmatrix} \operatorname{Re}\mathcal{S}f & -\operatorname{Im}\mathcal{S}f \\ -\operatorname{Im}\mathcal{S}f & -\operatorname{Re}\mathcal{S}f \end{pmatrix},$$

where $\mathcal{S}f$ is the classical Schwarzian derivative of f. In this way we identify $B_{\mathbf{g}_0}(\log|f'|)$ with $\mathcal{S}f$.

Next, if $f = h+\overline{g}$ is a harmonic mapping of $\mathbb{D}$ and $\sigma = \log(|h'|+|g'|)$ is the conformal factor associated with the lift $\widetilde{f}$, we put

$$\mathcal{S}f = \mathcal{S}_{\mathbf{g}_0}\widetilde{f} = B_{\mathbf{g}_0}(\sigma)\,.$$

Calculating this out and making the identification of the generalized Schwarzian with a complex number produces

$$B_{\mathbf{g}_0}(\sigma) = 2\big(\sigma_{zz} - \sigma_z^2\big)\,,$$

which is the definition of $\mathcal{S}f$ given in (1.6).

We need two more general facts. First, if we change a metric $\mathbf{g}$ conformally to $\widehat{\mathbf{g}} = e^{2\rho}\mathbf{g}$ then the tensor $B_{\mathbf{g}}(\psi)$ changes according to

$$B_{\mathbf{g}}(\rho+\psi) = B_{\mathbf{g}}(\rho) + B_{\widehat{\mathbf{g}}}(\psi)\,.$$

This is actually a generalization of the chain rule (1.1) for the Schwarzian. An equivalent formulation is

$$B_{\widehat{\mathbf{g}}}(\psi-\rho) = B_{\mathbf{g}}(\psi) - B_{\mathbf{g}}(\rho)\,, \tag{4.2}$$

which is what we will need in the next section.

Second, just as the linear differential equation $w'' + pw = 0$ is associated with $Sf = 2p$, so is there a linear differential equation associated with the Schwarzian tensor. If $B_{\mathbf{g}}(\psi) = p$, where p is a symmetric, traceless 2-tensor, then $\phi = e^{-\psi}\phi$ satisfies

$$\mathrm{Hess}_{\mathbf{g}}(\phi) + \phi p = \frac{1}{2}(\Delta_{\mathbf{g}}\phi)\mathbf{g}\,. \tag{4.3}$$

Although, the setting is more general, the substitution $\phi = e^{-\psi}$ is suggested by the classical case; regard $\mathcal{S}f = 2p$ as a Riccati equation $u' - (1/2)u^2 = 2p$ for $u = f''/f'$. With $v = \log f'$ and $w = e^{-v/2} = (f')^{-1/2}$ one finds that w satisfies $w'' + pw = 0$. See [21].

Finally, a comment about convexity. Let $\gamma(s)$ be an arc-length parametrized geodesic for the metric $\mathbf{g}$. Evaluating both sides of the Equation (4.3) at the pair $(\gamma'(s), \gamma'(s))$ gives the scalar equation

$$\frac{d^2}{ds^2}\phi(\gamma(s)) + \phi(\gamma(s))p\big(\gamma'(s), \gamma'(s)\big) = \frac{1}{2}\Delta_{\mathbf{g}}\phi(\gamma(s))\,;$$

this uses $\mathbf{g}(\gamma'(s), \gamma'(s)) = 1$. If

$$\frac{1}{2}\Delta_{\mathbf{g}}\phi(\gamma(s)) - \phi(\gamma(s))p\big(\gamma'(s), \gamma'(s)\big) \geq 0$$

for all geodesics then ϕ is convex relative to the metric $\mathbf{g}$. Without evaluating on a pair of vector fields, the condition for a function ϕ to be convex can be written as

$$\mathrm{Hess}_{\mathbf{g}}\,\phi \geq \alpha\mathbf{g}\,,$$

where α is a nonnegative function. We will find that an upper bound for $\mathcal{S}_{\mathbf{g}}f$ coming from the univalence criterion (1.7) leads via (4.2) and (4.3) to just such a positive lower bound for the Hessian of an associated function.

5. Univalence criteria and convexity

Convexity enters the picture by relating upper bounds on the Schwarzian tensor, in the guise of the Schwarzian of a harmonic mapping, to lower bounds on the Hessian of an associated function. There are two aspects to this. The first is to identify the appropriate conformal metric with respect to which the computations are made, and we do this now.

We recall the function

$$\Phi(x) = \int_0^x u_0(t)^{-2}\,dt\,, \qquad -1 < x < 1\,,$$

defined in (2.3), where u_0 is a positive solution of

$$u'' + pu = 0, \quad u(0) = 1, u'(0) = 0\,,$$

when p is a Nehari function. Recall also that Φ is odd with $\Phi(0) = 0$, $\Phi'(0) = 1$ and $\Phi''(0) = 0$.

We use Φ to form the radial conformal metric $\mathbf{g} = \Phi'(|z|)^2|dz|^2$ on $\mathbb{D}$. It is straightforward to express the curvature as

$$K_{\mathbf{g}}(z) = -2\big|\Phi'(r)\big|^{-2}(A(r) + p(r))\,, \qquad r = |z|\,, \tag{5.1}$$

where

$$A(r) = \frac{1}{4}\left(\frac{\Phi''(r)}{\Phi'(r)}\right)^2 + \frac{1}{2r}\frac{\Phi''(r)}{\Phi'(r)}\,; \tag{5.2}$$

see [8]. Note also that $A(r)$ is continuous at 0 with $A(0) = p(0)$, and that the curvature is *negative.* In particular,

$$|K_{\mathbf{g}}(z)| = 2\left|\Phi'(r)\right|^{-2}(A(r) + p(r))\,. \tag{5.3}$$

Appealing to the results in [8] we can assume that the metric $\Phi'(|z|)|dz|$ is complete, which means precisely that $\Phi(1) = \infty$. To elaborate, if $\Phi(1) < \infty$ then, as is shown in [8], there is a maximum value $t_0 > 1$ such that $t_0 p(x)$ remains a Nehari function and such that the corresponding extremal Φ_{t_0} has $\Phi_{t_0}(1) = \infty$. Since any condition of the form $|\mathcal{S}f| + \cdots \le 2p$ implies trivially that $|\mathcal{S}f| + \cdots \le 2t_0 p$ one may take $\Phi(1) = \infty$ at the outset. We make this assumption.

Geometric consequences of completeness of the metric are that any two points in $\mathbb{D}$ can be joined by a geodesic for $\mathbf{g}$, and that any geodesic can be extended indefinitely. An analytic consequence of completeness, following from $\Phi(1) = \infty$, is

$$p(x) \le A(x)\,. \tag{5.4}$$

This was shown in [8].

In Theorem 1.3 we have already seen Φ play the role of an extremal function for the univalence criterion. Our results on distortion and boundary behavior depend on Φ defining, as above, an 'extremal metric' for the criterion. In the planar, analytic case a detailed study was carried out in [8] and [9].

The second aspect of our analysis is captured in the following convexity result.

Theorem 5.1. *Let $f = h + \overline{g}$ be a harmonic mapping satisfying the conditions of Theorem 1.1, and in particular* (1.7),

$$|\mathcal{S}f(z)| + e^{2\sigma(z)}\left|K\left(\widetilde{f}(z)\right)\right| \le 2p(|z|) = S\Phi(|z|)\,.$$

Let $\varphi(z) = \log \Phi'(|z|)$ and define

$$u(z) = e^{(\varphi(z) - \sigma(z))/2} = \sqrt{\frac{\Phi'(|z|)}{\left|h'(z)\right| + \left|g'(z)\right|}}\,. \tag{5.5}$$

Then

$$\mathrm{Hess}_{\mathbf{g}}(u) \ge \frac{1}{4}u^{-3}|K|\mathbf{g}\,, \tag{5.6}$$

where $\mathbf{g}$ is the conformal metric $\Phi'(|z|)^2|dz|^2$. In particular, u is a convex function relative to $\mathbf{g}$.

Thus, we see that convexity obtains not for the conformal factor of $\widetilde{f}$, or for its square root, but for the square root of the *ratio* of the conformal factors of $\widetilde{f}$ and the extremal mapping for the univalence criterion.

The work of the present section is to establish Theorem 5.1. The inequality (5.4) is crucial, and to highlight this aspect of the proof we split off the following calculation as a separate lemma.

Lemma 5.2. *If f satisfies* (1.7) *then*

$$||B_{\mathbf{g}}(\sigma-\varphi)(z)||_{\mathbf{g}}+e^{2(\sigma-\varphi)(z)}\left|K\left(\widetilde{f}(z)\right)\right| \le \frac{1}{2}|K_{\mathbf{g}}(z)|\,. \tag{5.7}$$

Proof. Using the variant (4.2) of the addition formula for the Schwarzian tensor we have

$$B_{\mathbf{g}}(\sigma-\varphi)=B_{\mathbf{g}_0}(\sigma)-B_{\mathbf{g}_0}(\varphi)\,,$$

where $\mathbf{g}_0$ is the Euclidean metric. Now $\mathbf{g}=e^{2\varphi}\mathbf{g_0}$, and so the norm scales according to

$$e^{2\varphi(z)}||B_{\mathbf{g}}(\sigma-\varphi)(z)||_{\mathbf{g}}=||B_{\mathbf{g}}(\sigma-\varphi)(z)||_{\mathbf{g_0}}\,,$$

whence by the preceding equation

$$e^{2\varphi(z)}||B_{\mathbf{g}}(\sigma-\varphi)(z)||_{\mathbf{g}}=||B_{\mathbf{g}_0}(\sigma)(z)-B_{\mathbf{g}_0}(\varphi)(z)||_{\mathbf{g}_0}\,.$$

Finally,

$$||B_{\mathbf{g}_0}(\sigma)(z)-B_{\mathbf{g}_0}(\varphi)(z)||_{\mathbf{g}_0}=|B_{\mathbf{g}_0}(\sigma)(z)-B_{\mathbf{g}_0}(\varphi)(z)|\,,$$

which comes from identifying the tensor $B_{\mathbf{g}_0}(\sigma)(z)-B_{\mathbf{g}_0}(\varphi)(z)$ with the corresponding complex number, as explained in the previous section. A calculation then shows that the right-hand side can be expressed as

$$|B_{\mathbf{g}_0}(\sigma)(z)-B_{\mathbf{g}_0}(\varphi)(z)|=\left|\zeta^2\mathcal{S}f(z)+A(r)-p(r)\right|,\qquad r=|z|,\ \zeta=z/r\,,$$

where $A(r)$ is defined by (5.2).

Appealing now to (5.3), we see that proving (5.7) is equivalent to proving

$$\left|\zeta^2\mathcal{S}f(z)+A(r)-p(r)\right|+e^{2\sigma(z)}\left|K\left(\widetilde{f}(z)\right)\right|\le A(r)+p(r)\,.$$

But in view of the inequality (5.4) and the hypothesis (1.7), we now have

$$\begin{aligned}\left|\zeta^2\mathcal{S}f(z)+A(r)-p(r)\right|+e^{2\sigma(z)}\left|K\left(\widetilde{f}(z)\right)\right| &\le \left|\zeta^2\mathcal{S}f(z)\right|+|A(r)-p(r)|+e^{2\sigma(z)}\left|K\left(\widetilde{f}(z)\right)\right|\\ &= |\mathcal{S}f(z)|+e^{2\sigma(z)}\left|K\left(\widetilde{f}(z)\right)\right|+A(r)-p(r)\\ &\le A(r)+p(r)\,.\end{aligned}$$

□

Proof of Theorem 5.1. With $u=e^{(\varphi-\sigma)/2}$ and $v=u^2$ we find using (4.3) that

$$\mathrm{Hess}_{\mathbf{g}}(v)+vB_{\mathbf{g}}(\sigma-\varphi)=\frac{1}{2}(\Delta_{\mathbf{g}}v)\mathbf{g}\,.$$

Also $\Delta_{\mathbf{g}}=e^{-2\varphi}\Delta$, where Δ is the Euclidean Laplacian, thus

$$\begin{aligned}\Delta_{\mathbf{g}}v &= v\Delta_{\mathbf{g}}(\log v)+\frac{1}{v}||\,\mathrm{grad}_{\mathbf{g}}\,v||_{\mathbf{g}}^2\\ &= ve^{-2\varphi}\Delta(\varphi-\sigma)+\frac{1}{v}||\,\mathrm{grad}_{\mathbf{g}}\,v||_{\mathbf{g}}^2\,.\end{aligned}$$

Now using $K_{\mathbf{g}}=-e^{-2\varphi}\Delta\varphi$, $K=-e^{-2\sigma}\Delta\sigma$, and the fact that both curvatures are negative, we can rewrite this as

$$\Delta_{\mathbf{g}}v=v|K_{\mathbf{g}}|-ve^{2(\sigma-\varphi)}|K|+\frac{1}{v}||\,\mathrm{grad}_{\mathbf{g}}\,v||_{\mathbf{g}}^2\,,$$

and hence

$$\mathrm{Hess}_{\mathbf{g}}(v) = -vB_{\mathbf{g}}(\sigma-\varphi) + \frac{v}{2}\big(|K_{\mathbf{g}}| - e^{2(\sigma-\varphi)}|K|\big)\mathbf{g} + \frac{1}{2v}||\,\mathrm{grad}_{\mathbf{g}}\, v||^2_{\mathbf{g}}\,\mathbf{g}\,.$$

Therefore, because of the lemma,

$$\mathrm{Hess}_{\mathbf{g}}(v) \geq \left(\frac{v}{2}e^{2(\sigma-\varphi)}|K| + \frac{1}{2v}||\,\mathrm{grad}_{\mathbf{g}}\, v||^2_{\mathbf{g}}\right)\mathbf{g}\,. \tag{5.8}$$

On the other hand, since $v = u^2$,

$$\mathrm{Hess}_{\mathbf{g}}(v) = 2u\,\mathrm{Hess}_{\mathbf{g}}(u) + 2\,du \otimes du\,,$$

and

$$\frac{1}{2v}||\,\mathrm{grad}_{\mathbf{g}}\, v||^2_{\mathbf{g}} = 2||\,\mathrm{grad}_{\mathbf{g}}\, u||^2_{\mathbf{g}}\,.$$

We finally deduce from (5.8) that

$$2u\,\mathrm{Hess}_{\mathbf{g}}(u) + 2\,du \otimes du \geq \frac{v}{2}e^{2(\sigma-\varphi)}|K|\mathbf{g} + 2||\,\mathrm{grad}_{\mathbf{g}}\, u||^2_{\mathbf{g}}\,\mathbf{g}\,,$$

and since $du \otimes du$ is at most the norm-squared of the gradient,

$$2u\mathrm{Hess}_{\mathbf{g}}(u) \geq \frac{v}{2}e^{2(\sigma-\varphi)}|K|\mathbf{g}\,.$$

This is equivalent to (5.6). □

6. Critical points, distortion, and boundary behavior

We continue to assume that the harmonic mapping $f = h + \bar{g}$ satisfies the univalence criterion (1.7),

$$|\mathcal{S}f(z)| + e^{2\sigma(z)}\left|K\left(\widetilde{f}(z)\right)\right| \leq 2p(|z|) = \mathcal{S}\Phi(|z|)\,.$$

We also continue to work with the metric $\mathbf{g} = \Phi'(|z|)^2|dz|^2$ on $\mathbb{D}$.

In this section we will use the convexity of the function u defined in (5.5) in Theorem 5.1 to derive upper bounds on $e^{\sigma(z)} = |h'(z)|+|g'(z)|$. This is the key to obtaining continuous extensions of f and $\widetilde{f}$ to $\partial\mathbb{D}$ as stated in Theorem 1.2. The analysis leads to a more refined understanding of the phenomenon than given by the short, straightforward assertion of the theorem, but distinguishing special cases makes it difficult to collect all the results in a single statement.

An important issue is the number of critical points the function u can have, specifically when that number does not exceed one. This separates analytic maps, where we can appeal to earlier work, from harmonic maps, where the results of the preceding section will be applied. The distinction is made on the basis of the following lemma.

Lemma 6.1. *If u has more than one critical point then $\widetilde{f}(\mathbb{D})$ is a planar minimal surface.*

Proof. Suppose z_1 and z_2 are critical points of u. Then, because u is convex, $u(z_1)$ and $u(z_2)$ are absolute minima, and so is every point on the geodesic segment γ joining z_1 and z_2 in $\mathbb{D}$.

Hence, $\text{Hess}_{\mathbf{g}}(u)(\gamma', \gamma') = 0$, and then Theorem 5.1 implies that $|K| \equiv 0$ along $\widetilde{\gamma} = \widetilde{f}(\gamma)$. On the other hand, we know from (1.5) that

$$|K| = \frac{4|q'|^2}{|h'|^2(1+|q|^2)^4},$$

hence, $q' = 0$ along γ, so q' is identically 0. This proves the lemma. □

Remark. The proof shows a bit more than stated in the lemma, namely that the surface Σ will reduce to a plane provided $\text{Hess}_{\mathbf{g}}(u)(\gamma', \gamma') = 0$ along any geodesic segment γ. In particular, this will be the case if u is constant along γ.

Thus, in the case of multiple critical points the lifted map $\widetilde{f}$ can be considered as a *holomorphic* mapping into a tilted (complex) plane, and it satisfies the classical Nehari condition

$$\left|\mathcal{S}\widetilde{f}(z)\right| \le 2p(|z|).$$

A fairly complete treatment of such mappings, specifically continuous extension to the boundary, extremal functions (and homeomorphic or quasiconformal extensions to $\mathbb{C}$) can be found in [14] and [7]. Briefly, the boundary behavior of $\widetilde{f}$ is of the same character as that of the extremal function Φ near $x = 1$ in the spherical metric (recall that $\Phi(1) = \infty$), a phenomenon we will find to hold as well when u has one or no critical points. When $p(x) = (1 - x^2)^{-2}$ the extremal is a logarithm and $\widetilde{f}$ has a logarithmic modulus of continuity. For all other choices of Nehari functions the extension is Hölder continuous.

We next consider the case when u has exactly one critical point. Under that condition, the following lemma is the promised upper bound for $|h'| + |g'|$.

Lemma 6.2. *If u has a unique critical point then there exist constants $a > 0$, b, and r_0 $(0 < r_0 < 1)$ such that*

$$\left|h'(z)\right| + \left|g'(z)\right| \le \frac{\Phi'(|z|)}{(a\Phi(|z|) + b)^2}, \qquad r_0 < |z| < 1. \tag{6.1}$$

Proof. Let z_0 be the unique critical point of u. Let $\gamma(s)$ be an arclength parametrized geodesic in the metric $\mathbf{g}$ starting at z_0 in a given direction. Let $v(s) = u(\gamma(s))$. Because the critical point is unique, it follows that $v'(s) > 0$ for all $s > 0$, and hence that there exist an $s_0 > 0$ and an $a > 0$ such that $v'(s) > a$ for all $s > s_0$. In turn, $v(s) > as + b$ for some constant b and all $s > s_0$, and from compactness we can conclude that the constants s_0, a, b in this estimate can be chosen uniformly, independent of the direction of the geodesic starting at z_0. In other words,

$$u(z) \ge a d_{\mathbf{g}}(z, z_0) + b$$

for all z with $d_{\mathbf{g}}(z, z_0) > s_0$, where $d_{\mathbf{g}}$ denotes distance in the metric $\mathbf{g}$. Then by renaming the constant b, and with a suitable r_0, we will have

$$u(z) \ge a d_{\mathbf{g}}(z, 0) + b$$

for all z with $|z| > r_0$. Since $d_{\mathbf{g}}(z, 0) = \Phi(|z|)$ the lemma follows. □

We view Lemma 6.2 as a distortion theorem for harmonic mappings satisfying the univalence criterion, and it is the estimate (6.1) that will allow us to obtain a continuous extension to the closed disk for the lift $\widetilde{f}$ and for the harmonic mapping f. The modulus of continuity of the extension depends on particular properties of the function Φ, and ranges from a logarithmic modulus of continuity, to Hölder and to Lipschitz continuity.

We begin by observing that since the function $(1-x^2)^2 p(x)$ is positive and nonincreasing on $[0, 1)$ we can form the limit

$$\lambda = \lim_{x\to 1^-} \left(1-x^2\right)^2 p(x)\,.$$

It was shown in [8] that $\lambda \le 1$ and that $\lambda = 1$ if and only if $p(x) = (1-x^2)^{-2}$. Consider first this case, when $p(x) = (1-x^2)^{-2}$. Then

$$\Phi(x) = \frac{1}{2}\log\frac{1+x}{1-x}\,,$$

and (6.1) amounts to

$$\left|h'(z)\right| + \left|g'(z)\right| \le \frac{1}{\left(1-|z|^2\right)\left(\dfrac{a}{2}\log\dfrac{1+|z|}{1-|z|}+b\right)^2}\,, \qquad r_0 < |z| < 1\,. \tag{6.2}$$

From here, to show that $\widetilde{f}$ extends continuously to the closed disk, we follow the argument in [14] and integrate along a hyperbolic geodesic. Let ϱ be the hyperbolic segment joining two points z_1 and z_2 in $\mathbb{D}$. Then ϱ has Euclidean length $\ell \le \frac{\pi}{2}|z_1-z_2|$ and $\min(s, \ell-s) \le \frac{\pi}{2}(1-|z|)$ for each $z \in \varrho$, where s is the Euclidean arclength of the part of ϱ between z_1 and z. Suppose that z_1 and z_2 are such that ϱ is contained in the annulus $r_0 < |z| < 1$. The distance $|\widetilde{f}(z_1) - \widetilde{f}(z_2)|$ in $\mathbb{R}^3$ is less than the metric distance between $\widetilde{f}(z_1)$ and $\widetilde{f}(z_2)$ on the surface Σ, and this in turn is less than the length of ϱ in the metric $(|h'|+|g'|)|dz|$ on $\mathbb{D}$. Thus, we may use (6.1) and (6.2) and write

$$\begin{aligned}
\left|\widetilde{f}(z_1) - \widetilde{f}(z_2)\right| &\le \int_\varrho \left(\left|h'(z)\right| + \left|g'(z)\right|\right)|dz| \le \int_\varrho \frac{|dz|}{\left(1-|z|^2\right)\left(\frac{a}{2}\log\frac{1+|z|}{1-|z|}+b\right)^2} \\
&\le C\int_0^{\ell/2} \frac{ds}{(1-s)\left(\frac{a}{2}\log\frac{1}{1-s}+b\right)^2}\,,
\end{aligned}$$

for some constant C independent of z_1 and z_2.

Integration, together with the bound $\ell \le \frac{\pi}{2}|z_1 - z_2|$, yields

$$\left|\widetilde{f}(z_1) - \widetilde{f}(z_2)\right| \le \frac{C'}{\log\dfrac{1}{|z_1-z_2|}}\,,$$

for another constant C'. This implies that $\widetilde{f}$ has an extension to $\overline{\mathbb{D}}$ that is uniformly continuous. Since

$$|f(z_1) - f(z_2)| \le \left|\widetilde{f}(z_1) - \widetilde{f}(z_2)\right|, \tag{6.3}$$

the same is true for the harmonic mapping f. In all the cases that we consider, it is simply the inequality (6.3) that is used to obtain a continuous extension for f from one for $\widetilde{f}$.

Suppose now that $\lambda < 1$. It was shown in [8] that

$$\lim_{x\to 1^-}\big(1-x^2\big)\frac{\Phi''(x)}{\Phi'(x)} = 2\big(1+\sqrt{1-\lambda}\big) = 2\mu\,. \tag{6.4}$$

Note that $1 < \mu \le 2$. It follows from (6.4) that for any $\varepsilon > 0$ there exists $x_0 \in (0, 1)$ such that

$$\frac{\mu-\varepsilon}{1-x} \le \frac{\Phi''(x)}{\Phi'(x)} \le \frac{\mu+\varepsilon}{1-x}\,, \qquad x > x_0\,.$$

This implies that

$$\frac{1}{(1-x)^{\mu-\varepsilon}} \le \Phi'(x) \le \frac{1}{(1-x)^{\mu+\varepsilon}}\,, \qquad x > x_0\,,$$

so that

$$\frac{\Phi'(x)}{(a\Phi(x)+b)^2} \le \frac{C}{(1-x)^{\alpha+3\varepsilon}}\,, \tag{6.5}$$

where $\alpha = 2-\mu = 1-\sqrt{1-\lambda}$ and C depends on a, b, and the values of Φ at x_0. The estimate in (6.5) together with the technique of integrating along a hyperbolic segment implies now that

$$\big|\widetilde{f}(z_1) - \widetilde{f}(z_2)\big| \le C|z_1 - z_2|^{1-\alpha-3\varepsilon} = C|z_1 - z_2|^{\sqrt{1-\lambda}-3\varepsilon}\,,$$

for all points z_1, z_2 whose joining hyperbolic geodesic segment is contained in the annulus $\max\{r_0, x_0\} < |z| < 1$. This shows that $\widetilde{f}$, and hence f, admits a continuous extension to the closed disk, with at least a Hölder modulus of continuity.

Observe that the left-hand side of (6.5) is the derivative of a Möbius transformation of Φ. Thus, the modulus of continuity of $\widetilde{f}$, as derived from this bound, is, in essence, that of the extremal at $x = 1$ in the spherical metric.

We also remark that if additionally it is known that $x = 1$ is a regular singular point of the differential equation $u'' + pu = 0$, then from the analysis of the Frobenius solutions at $x = 1$ it follows that

$$\Phi'(x) \sim \frac{1}{(1-x)^{\mu}}\,, \qquad x \to 1\,.$$

This then provides for a Hölder continuous extension with the 'best' exponent when $\lambda > 0$, and a Lipschitz continuous extension when $\lambda = 0$.

Recall that the preceding arguments were carried out under the umbrella of Lemma 6.2, when u has a unique critical point. In this case, the extension to the closed disk is continuous (or more) in the Euclidean metric. We finally treat the case when u has no critical points, and here the situation is somewhat different.

A fairly straightforward argument gives half a result, so to speak. Suppose that u has no critical points and let $u_\theta(s) = u(r(s)e^{i\theta})$, where $r(s)$ is the arclength parametrization of the radius $[0, 1)$ in the metric $\mathbf{g}$. By assumption, the gradient of u at the origin does not vanish, and hence $u'_\theta(0) > 0$ for all arguments θ in an open half-plane; to be specific, say $u'_\theta(0) > 0$ for $0 < \theta < \pi$. Furthermore, because u cannot be constant along any geodesic unless the surface Σ reduces to a plane, it follows that $u'_\theta(s) > 0$ for $s > 0$ and $\theta = 0$ or $\theta = \pi$. Now again by compactness we see that there is an s_0 such that $u'_\theta(s) > a > 0$ for $s \ge s_0$ and all $0 \le \theta \le \pi$. From here we can pick up the proof of Lemma 6.2 and deduce that (6.1) holds for all $|z| > r_0$ with $\mathrm{Im}\{z\} \geq 0$. To reiterate, this then provides a continuous extension of $\widetilde{f}$ and of f to the upper half of $\overline{\mathbb{D}}$.

To get beyond this half-disk result we will show that for any radius $[0, e^{i\theta_0})$ there exists a Möbius transformation T of $\mathbb{R}^3$ such that the conformal factor associated with the conformal lift $T \circ \widetilde{f}$ satisfies a version of (6.1) in an angular sector $|\theta - \theta_0| < \delta$ about the radius—why we need the extra Möbius transformation will emerge presently. Thus, $T \circ \widetilde{f}$ will exhibit the appropriate continuous extension. It is because we have to allow for a shift of the range by a Möbius transformation of $\mathbb{R}^3$ that the summary result on boundary extensions, Theorem 1.2, is stated to assert that $\widetilde{f}$ and f have extensions to the closed disk that are continuous in the *spherical* metric.

Let T be a Möbius transformation of $\mathbb{R}^3$. Since $T \circ \widetilde{f}$, though conformal, may not be the lift of a harmonic mapping we do not have the basic convexity result Theorem 5.1 as a starting point. We shall first find a substitute that holds along a radius.

Let e^τ be the conformal factor associated with $T \circ \widetilde{f}$, that is, $e^{\tau(z)} = |DT(\widetilde{f}(z))|e^{\sigma(z)}$, and for fixed θ let $\tau_\theta(r) = \tau(re^{i\theta})$. Along each radius we have the following.

Lemma 6.3. *Let $r = r(s)$ be the arclength parametrization of $[0, 1)$ in the metric $\mathbf{g} = \Phi'(|z|)^2|dz|^2$. Then the function*

$$\omega(s) = \sqrt{\frac{\Phi'(r(s))}{e^{\tau_\theta(r(s))}}} \tag{6.6}$$

is a convex function of s.

Before giving the proof we note that it is easy to identify the arclength parametrization of a radius in the metric $\mathbf{g}$. Since Φ is radial the length s of the radius $[0, re^{i\theta}]$ is

$$s = \int_0^r \Phi'(t)\, dt = \Phi(r)\,.$$

Thus, if Ψ denotes the inverse of Φ, then $r(s) = \Psi(s)$ in the notation of the lemma.

Proof. Let $w(r) = \Phi'(r)^{-1/2}$, so that $w'' + pw = 0$, and let $g(r) = (T \circ \widetilde{f})(re^{i\theta})$ and note that $|g'(r)| = e^{\tau_\theta(r)}$. A straightforward calculation shows that the function $v(r) = |g'(r)|^{-1/2}$ satisfies the equation $v'' + qv = 0$ with

$$q = \frac{1}{2}S_1 g - \frac{1}{4}\left(\frac{|g''|^2}{|g'|^2} - \frac{\langle g'', g'\rangle^2}{|g'|^4}\right),$$

where S_1 is Ahlfors' Schwarzian (2.1). The quantity in parentheses is nonnegative, and by Möbius invariance, $S_1 g(r) = S_1 \widetilde{f}(re^{i\theta})$. Since $S_1 \widetilde{f}(re^{i\theta}) \leq 2p(re^{i\theta})$ by assumption, we conclude that $q \leq p$.

Introducing $\Psi = \Phi^{-1}$ as above, we write

$$\omega(s) = \sqrt{\frac{\Phi'(r(s))}{e^{\tau_\theta(r(s))}}} = \frac{v(\Psi(s))}{w(\Psi(s))}\,,$$

and now one can check that

$$\omega'' = (p - q)w^4\omega\,,$$

where differentiation is with respect to s and the quantity $(p-q)w^4$ is to be evaluated at $\Psi(s)$. Since $p \geq q$, it follows that ω is a convex function of s. □

To obtain an estimate of the type (6.1) along a given radius $[0, e^{i\theta})$ it will suffice to show that the function ω in (6.6) has $\omega'(0) > 0$. Indeed, if this derivative is positive then, once more by convexity, $\omega(s) \geq as + b$ for some constant b and some positive constant a. With this,

$$e^{\tau_\theta(r)} \leq \frac{\Phi'(r)}{(a\Phi(r)+b)^2}\,, \tag{6.7}$$

corresponding to (6.1). Then, if for a given angle θ_0 the derivative of ω at zero is positive, by continuity it will remain positive for all angles θ close to θ_0. Therefore (6.7) will hold uniformly in an angular sector, leading to a continuous extension of $\widetilde{f}$ and f to the part of the boundary within the sector, just as before.

To finish the argument, we thus need to find a Möbius transformation T of $\mathbb{R}^3$ so that, for a given angle θ_0, the function ω has positive derivative at zero. Using $\Phi''(0) = 0$ we are therefore required to have $\tau'_{\theta_0}(0) < 0$. After a translation, a rotation and a dilation, we can assume that the curve $\widetilde{f}_{\theta_0}(r) = \widetilde{f}(re^{i\theta_0})$ has $\widetilde{f}_{\theta_0}(0) = 0$, $\widetilde{f}'_{\theta_0}(0) = (1,0,0)$ and $\widetilde{f}''_{\theta_0}(0) = (\alpha,\beta,0)$. Let T be the extension as a Möbius transformation of $\mathbb{R}^3$ of the complex Möbius map

$$z \mapsto \frac{z}{1+cz}\,,$$

where $c = (1+\alpha)/2$ and we identify $z = x+iy$ with the point $(x, y, 0)$. Since up to order 2 the curve $\widetilde{f}_{\theta_0}(r)$ lies in the (x, y)-plane, for the purposes of our calculations, which involve derivatives of order 2 at most, we may replace the curve $(T \circ \widetilde{f}_{\theta_0})(r)$ with the curve

$$w(r) = \frac{z(r)}{1+cz(r)}\,,$$

where $z(r)$ satisfies $z(0) = 0$, $z'(0) = 1$, $z''(0) = \alpha + i\beta$. Up to an error of order $O(r^2)$ we have that $\tau_{\theta_0}(r) = \log|w'(r)| = \log|z'(r)| - 2\log|1+cz(r)|$, from which $\tau'_{\theta_0}(0) = \alpha - 2c = -1 < 0$, as desired.

Remark. The Möbius transformation T may send some point on the surface Σ to the point at infinity, but such a point cannot lie in the image of $\widetilde{f}_{\theta_0}(r)$. Indeed, once we have ensured that the function ω in Lemma 6.3 has $\omega'(0) > 0$, then the estimate (6.7) will imply that the curve $(T \circ \widetilde{f}_{\theta_0})(r)$ has finite length, and it is therefore impossible for it to reach the point at infinity.

The arguments in this section, supported by the results of the previous two sections, have proved Theorem 1.2 as stated, that f and $\widetilde{f}$ have spherically continuous extensions to $\overline{\mathbb{D}}$. Much more detail has been obtained *en route,* and we conclude with an expanded, if admittedly under-specified version of the theorem that we hope helpfully captures the main points.

Theorem 6.4. *Suppose f satisfies the univalence criterion*

$$|\mathcal{S}f(z)| + e^{2\sigma(z)}\left|K\big(\widetilde{f}(z)\big)\right| \leq 2p(|z|) = \mathcal{S}\Phi(|z|)\,,$$

with extremal function Φ, and let $\lambda = \lim_{x\to1^-}(1-x^2)^2 p(x)$. Then f and $\widetilde{f}$ have extensions to $\overline{\mathbb{D}}$ that are continuous with respect to the spherical metric. The modulus of continuity of each is of the same type as that of $\Phi(x)$ near $x = 1$ in the spherical metric. If $\lambda = 1$ it is logarithmic. If $\lambda < 1$ it is Hölder with an exponent that depends on λ.

7. The catenoid and extremal lifts

The principal work of this section is to consider some examples that show our results are sharp. The constructions are based on what we know from the analytic case, and on one of the earliest minimal surfaces to be studied, the catenoid. According to Theorem 1.3, the minimal surface corresponding to the lift of an extremal mapping must contain a Euclidean circle or line as a circle of curvature. Catenoids enter naturally into the discussion of extremal lifts because they are the *unique* minimal surfaces containing a Euclidean circle as a line of curvature, as we shall now show.

Lemma 7.1. *If a minimal surface Σ contains a part of a Euclidean circle or line as a line of curvature, then Σ is contained in a catenoid or a plane.*

Proof. From the theory of minimal surfaces we will need a uniqueness result associated with the *Björling problem* of finding a minimal surface with a prescribed normal strip. This may be stated as follows. Let $I \subset \mathbb{R}$ be an open interval. A real-analytic strip $S = \{(\mathbf{c}(t), \mathbf{n}(t)) : t \in I\}$ in $\mathbb{R}^3$ consists of a real-analytic curve $\mathbf{c} : I \to \mathbb{R}^3$ with $\mathbf{c}'(t) \neq 0$ and a real-analytic vector field $\mathbf{n} : I \to \mathbb{R}^3$ along $\mathbf{c}$, with $|\mathbf{n}(t)| \equiv 1$ and $\langle \mathbf{c}'(t), \mathbf{n}(t)\rangle \equiv 0$. The problem is to find a parametrized minimal surface $\mathbf{X} : \Omega \to \mathbb{R}^3$ with $I \subset \Omega \subset \mathbb{R}^2$, such that $\mathbf{X}(x, 0) = \mathbf{c}(x)$, $\mathbf{N}(x, 0) = \mathbf{n}(x)$ for $x \in I$, where $\mathbf{N}(x, y)$ is a unit normal vector field along the surface. The result we need is that for real-analytic data, Björling's problem admits exactly one solution; see [10], p. 121, where the solution is expressed in closed form in terms of the data defining the strip.

Suppose now that C is a Euclidean circle, part (or all) of which is a line of curvature of a minimal surface Σ. Let Π be the plane containing C, and let $\mathbf{N}$, $\mathbf{n}_0$ be, respectively, unit normal vectors to Σ and to Π. It follows from the classical theorem of Joachimstahl (see, for example, [11] p. 152) that the normal vector of a surface along a planar curve which is a line of curvature forms a constant angle with the normal to the plane of the curve. Therefore, in our case, $\mathbf{N}$ and $\mathbf{n}_0$ form a constant angle along C, say α.

Let Σ_0 be a catenoid and consider the situation above for Σ_0. All of the circles of revolution on Σ_0 are lines of curvature for Σ_0, and the angle between the normal to Σ_0 and the plane of any such circle decreases from $\pi/2$, for the circle around the waist of Σ_0, down toward 0 as the circles move out toward infinity. Choose one circle, C_0, where the angle is α. Via Euclidean similarities we can assume that C_0 and C coincide, and so both surfaces Σ and Σ_0 have the same unit normal vector fields along $C = C_0$.

Next let $\mathbf{X}$ and $\mathbf{X}_0$ be conformal parametrization of Σ and Σ_0, respectively, covering a part of the circle $C = C_0$, say C'. We arrange the parametrization of Σ_0 so that $\mathbf{X}_0^{-1}(C') = I$, an open interval. Because the preimage $c' = \mathbf{X}^{-1}(C')$ is a real analytic, simple curve, there is an invertible holomorphic map, h, of a neighborhood of I to a neighborhood of c' with $\mathbf{X}(h(t)) = \mathbf{X}_0(t)$, $t \in I$. It now follows from the uniqueness of the solution to Björling's problem that $\mathbf{X} = \mathbf{X}_0 \circ h^{-1}$ on a neighborhood of c'. Thus, Σ and Σ_0 coincide near C', and hence Σ is a portion of the catenoid. If Σ contains part of a Euclidean line as a line of curvature, instead of a circle, then a similar argument shows that Σ must be contained in a plane. □

With this result, we now have the following corollary of Theorem 1.3 on extremal maps.

Corollary 7.2. *Let f be an extremal mapping. Then $\widetilde{f}(\mathbb{D})$ is contained in a catenoid or a plane.*

We now proceed with examples. The case of an extremal lift mapping the disk into a plane is essentially the case of an analytic extremal, and examples there have been studied. More interesting for the present article are extremal mappings into a catenoid, where the analytic case can still serve as a guide.

Example 7.3. The choice $p(x) = \pi^2/4$ gives Nehari's univalence criterion (1.3) for analytic functions in the disk. In the analytic case an extremal mapping is

$$F(z) = \frac{2}{\pi} \tan\left(\frac{\pi}{2} z\right),$$

which maps $\mathbb{D}$ to a horizontal strip. For harmonic maps the criterion (1.7) becomes

$$|\mathcal{S} f(z)| + e^{2\sigma(z)} \left|K\left(\widetilde{f}(z)\right)\right| \leq \frac{\pi^2}{2}. \tag{7.1}$$

To show that the criterion is sharp we will work with the harmonic mapping

$$f(z) = h(z) + \overline{g(z)} = ce^{\pi z} + \frac{1}{c} e^{-\pi \bar{z}},$$

for a positive constant c to be chosen later. The lift $\widetilde{f}$ maps $\mathbb{D}$ into the catenoid parametrized by

$$\begin{aligned} U(x, y) &= \left(ce^{\pi x} + \frac{1}{c} e^{-\pi x}\right) \cos \pi y \\ V(x, y) &= \left(ce^{\pi x} + \frac{1}{c} e^{-\pi x}\right) \sin \pi y \\ W(x, y) &= 2\pi x \end{aligned}$$

with $z = x + iy$. The lift fails to be univalent at $\pm i$, with $\widetilde{f}(i) = \widetilde{f}(-i) = (-(c + \frac{1}{c}), 0, 0)$. In fact, the diameter $-1 \leq y \leq 1$ maps to the circle $U^2 + V^2 = (c + \frac{1}{c})^2$, $W = 0$ on the surface. This is one of the circles of revolution of the catenoid, and it is a line of curvature as guaranteed by Theorem 1.3.

To see what happens with (7.1), we find first that

$$e^{\sigma(z)} = \left|h'(z)\right| + \left|g'(z)\right| = \pi \left(ce^{\pi x} + \frac{1}{c} e^{-\pi x}\right),$$

and then for the Schwarzian,

$$\mathcal{S} f(z) = -\frac{\pi^2}{2} + 4\pi^4 e^{-2\sigma(z)}.$$

For the curvature term,

$$e^{2\sigma(z)} \left|K\left(\widetilde{f}(z)\right|\right. = \Delta\sigma(z) = 4\pi^4 e^{-2\sigma(z)}.$$

Therefore, (7.1) will be satisfied provided

$$\left|-\frac{\pi^2}{2} + 4\pi^4 e^{-2\sigma(z)}\right| + 4\pi^4 e^{-2\sigma(z)} \leq \frac{\pi^2}{2}.$$

This will be the case if $c > (1 + \sqrt{2})e^{\pi} = 55.866\ldots$, because then

$$\left|-\frac{\pi^2}{2} + 4\pi^4 e^{-2\sigma(z)}\right| = \frac{\pi^2}{2} - 4\pi^4 e^{-2\sigma(z)},$$

and for c in this range

$$|\mathcal{S}f(z)| + e^{2\sigma(z)}\big|K\big(\widetilde{f}(z)\big)\big| \equiv \frac{\pi^2}{2}\,.$$

By modifying this construction slightly we can also show that the constant $\pi^2/2$ is best possible. For the harmonic mapping take

$$f(z) = ce^{t\pi z} + \frac{1}{c}e^{-t\pi\bar{z}}\,, \quad t > 0\,.$$

Then

$$e^{\sigma(z)} = t\pi\left(ce^{t\pi x} + \frac{1}{c}e^{-t\pi x}\right),$$

and

$$\mathcal{S}f = -t^2\frac{\pi^2}{2} + 4t^4\pi^4e^{-2\sigma}\,,$$

while

$$e^{-2\sigma}|K| = \Delta\sigma = 4t^4\pi^4e^{-2\sigma}\,.$$

Therefore, if $c > 56$,

$$|\mathcal{S}f| + e^{-2\sigma}|K| = \left|-t^2\frac{\pi^2}{2} + 4t^4\pi^4e^{-2\sigma}\right| + 4t^4\pi^4e^{-2\sigma} = t^2\frac{\pi^2}{2}\,.$$

But as soon as $t > 1$ both the maps f and $\widetilde{f}$ fail to be univalent in $\mathbb{D}$.

Example 7.4. Portions of the catenoid also provide examples for other Nehari functions. We discuss a general procedure. Let p be a Nehari function that is the restriction to $(-1, 1)$ of an analytic function $p(z)$ in the disk with the property $|p(z)| \le p(|z|)$. Typical examples are $p(z) = (1 - z^2)^{-2}$ and $p(z) = 2(1 - z^2)^{-1}$. The extremal map in such a case, say F, can be normalized in the same way as the extremal Φ in (2.3), and is analytic, odd, and univalent in the disk and satisfies $SF(z) = 2p(z)$. The image $F(\mathbb{D})$ is a "parallel strip" like domain, symmetric with respect to both axis, and containing the entire real line; see [7]. Let

$$G(z) = \frac{cF(z) + i}{cF(z) - i}\,,$$

where $c > 0$ is to be chosen later and sufficiently small so that $i/c \notin F(\mathbb{D})$ (it can be shown that the map F is always bounded along the imaginary axis). The function G maps $\mathbb{D}$ onto a simply-connected domain containing the unit circle minus the point 1. Let

$$f(z) = h(z) + \overline{g(z)} = G(z) + \overline{G(-z)}\,.$$

Since F is odd,

$$f(z) = G(z) + \frac{1}{\overline{G(z)}}\,,$$

and it follows that the lift $\widetilde{f}$ parametrizes the catenoid with the unit circle $|G| = 1$ mapped onto the circle of symmetry of the catenoid. We also have

$$e^{\sigma} = \left|G'(z)\right| + \left|G'(-z)\right| = \frac{2c\left|F'(z)\right|}{|cF(z) - i|^2} + \frac{\left|2cF'(-z)\right|}{|cF(-z) - i|^2} = \frac{2c\left|F'(z)\right|}{|cF(z) - i|^2} + \frac{2c\left|F'(z)\right|}{|cF(z) + i|^2},$$

using again that F is odd. A somewhat tedious calculation shows that

$$\mathcal{S}f = 2(\sigma_{zz} - \sigma_z) = \mathcal{S}F - \frac{4c^2\left(1 + c^2\overline{F}^2\right)\left(F'\right)^2}{\left(1 + c^2F^2\right)\left(1 + c^2|F|^2\right)^2},$$

and

$$e^{2\sigma}|K| = \frac{4c^2\left|F'\right|^2}{\left(1 + c^2|F|^2\right)^2}.$$

Condition (1.7) now reads

$$\left|\mathcal{S}F(z) - \frac{4c^2\left(1 + c^2\overline{F(z)}^2\right)F'(z)^2}{\left(1 + c^2F(z)^2\right)\left(1 + c^2|F(z)|^2\right)^2}\right| + \frac{4c^2\left|F'(z)\right|^2}{\left(1 + c^2|F(z)|^2\right)^2} \le \mathcal{S}F(|z|). \tag{7.2}$$

Suppose, for example, we let $p(z) = (1 - z^2)^{-2}$, for which the extremal function is

$$F(z) = \frac{1}{2}\log\frac{1 + z}{1 - z}.$$

One has

$$F'(z) = \frac{1}{1 - z^2}, \quad \mathcal{S}F(z) = \frac{2}{\left(1 - z^2\right)^2},$$

and (7.2) becomes

$$\left|\frac{2}{\left(1 - z^2\right)^2} - \frac{4c^2\left(1 + c^2\overline{F(z)}^2\right)}{\left(1 - z^2\right)^2\left(1 + c^2F(z)^2\right)\left(1 + c^2|F(z)|^2\right)^2}\right| + \frac{4c^2}{\left|1 - z^2\right|^2\left(1 + c^2|F(z)|^2\right)^2} \le \frac{2}{\left(1 - |z|^2\right)^2},$$

which reduces to

$$\left|1 - \frac{2c^2\left(1 + c^2\overline{F(z)}^2\right)}{\left(1 + c^2F(z)^2\right)\left(1 + c^2|F(z)|^2\right)^2}\right| + \frac{2c^2}{\left(1 + c^2|F(z)|^2\right)^2} \le \frac{\left|1 - z^2\right|^2}{\left(1 - |z|^2\right)^2}. \tag{7.3}$$

We comment at once that equality holds here if z is real and if c is sufficiently small, for both sides of the inequality are then just 1. The task is to show that (7.3) holds for all $z \in \mathbb{D}$.

Let

$$\zeta = \frac{2c^2\left(1 + c^2\overline{F(z)}^2\right)}{\left(1 + c^2F(z)^2\right)\left(1 + c^2|F(z)|^2\right)^2}.$$

We establish the following estimates.

Lemma 7.5. *If c is small then there exist absolute constants A, B, C such that*

$$|1-\mathrm{Re}\{\zeta\}| \le 1-\frac{2c^2}{\left(1+c^2|F(z)|^2\right)^2}+Ac^4|\mathrm{Im}\{F(z)\}|^2\,, \tag{7.4}$$

$$|\mathrm{Im}\{\zeta\}| \le Bc^3|\mathrm{Im}\{F(z)\}|\,, \tag{7.5}$$

and

$$|1-\zeta| \le 1-\frac{2c^2}{\left(1+c^2|F(z)|^2\right)^2}+Cc^4|\mathrm{Im}\{F(z)\}|^2\,. \tag{7.6}$$

Proof. We write

$$\begin{aligned} 1-\mathrm{Re}\{\zeta\} &= 1-\frac{2c^2}{\left(1+c^2|F(z)|^2\right)^2}\mathrm{Re}\left\{\frac{1+c^2\overline{F(z)}^2}{1+c^2F(z)^2}\right\} \\ &= 1-\frac{2c^2}{\left(1+c^2|F(z)|^2\right)^2}+\frac{2c^2}{\left(1+c^2|F(z)|^2\right)^2}\left(1-\mathrm{Re}\left\{\frac{1+c^2\overline{F(z)}^2}{1+c^2F(z)^2}\right\}\right), \end{aligned}$$

which after some calculations gives

$$1-\mathrm{Re}\{\zeta\} = 1-\frac{2c^2}{\left(1+c^2|F(z)|^2\right)^2}+\frac{2c^4\left|cF(z)+c\overline{F(z)}\right|^2}{\left|1+c^2F(z)^2\right|^2\left(1+c^2|F(z)|^2\right)^2}\left|F(z)-\overline{F(z)}\right|^2.$$

The inequality (7.4) follows from this because the quantity

$$\frac{\left|cF(z)+c\overline{F(z)}\right|^2}{\left|1+c^2F(z)^2\right|^2\left(1+c^2|F(z)|^2\right)^2}$$

is uniformly bounded for small c.

To establish (7.5) we have

$$\begin{aligned} \mathrm{Im}\{\zeta\} &= \frac{2c^2}{\left(1+|F(z)|^2\right)^2}\mathrm{Im}\left\{\frac{1+c^2\overline{F(z)}^2}{1+c^2F(z)^2}\right\} \\ &= 2c^3\frac{\left(2+c^2F(z)^2+c^2\overline{F(z)}^2\right)\left(cF(z)+c\overline{F(z)}\right)}{\left|1+c^2F(z)^2\right|^2\left(1+|F(z)|^2\right)^2}\,\frac{\overline{F(z)}-F(z)}{2i}\,, \end{aligned}$$

from which (7.5) follows since

$$\frac{\left(2+c^2F(z)^2+c^2\overline{F(z)}^2\right)\left(cF(z)+c\overline{F(z)}\right)}{\left|1+c^2F(z)^2\right|^2\left(1+|F(z)|^2\right)^2}$$

is uniformly bounded for small c. Finally, (7.6) is a consequence of (7.4) and (7.5) because for $|\zeta|$ small, $\zeta=x+iy$, we have $|1-\zeta|\le|1-x|+2y^2$. □

With this lemma we can now obtain (7.3) via an analysis along the level sets of the function $|1-z^2|/(1-|z|^2)$. The set of points z where $|1-z^2|/(1-|z|^2)=\sqrt{1+t^2}$, $t>0$, corresponds to a pair of arcs of circles through ± 1 centered at $\pm i/t$ with radius $\sqrt{1+t^2}/t$. Because of the symmetry of (7.3) it suffices to consider the part of the upper arc, call it γ. The arc γ intersects the imaginary axis at $it/(1+\sqrt{1+t^2})$ and is mapped under F to the horizontal line $\operatorname{Im}\{F\} = \tan^{-1}(s)$, where $s = t/(1+\sqrt{1+t^2}) \le t$. Thus, $\operatorname{Im}\{F\} \le \tan^{-1} t \le t$. From (7.6) it follows that along γ

$$|1-\zeta| \le 1 - \frac{2c^2}{\left(1+c^2|F(z)|^2\right)^2} + Cc^3t^2\,,$$

so that the left-hand side of (7.3) is bounded above by $1 + Cc^4t^2 < 1 + t^2$ for c sufficiently small.

We have shown that the criterion

$$|\mathcal{S}f(z)| + e^{2\sigma(z)}\left|K\left(\widetilde{f}(z)\right)\right| \le \frac{2}{\left(1-|z|^2\right)^2}$$

is sharp. By adapting an example given by Hille [15] (which accompanied Nehari's original article) we can also show that the constant 2 in the numerator of the right-hand side is best possible. For this take

$$F(z) = \left(\frac{1+z}{1-z}\right)^{i\varepsilon}\,,$$

which is far from univalent in the unit disk if $\varepsilon > 0$; the value 1 is assumed infinitely often. Note that

$$F'(z) = \frac{2i\varepsilon}{1-z^2}F(z)\,, \qquad \mathcal{S}F(z) = \frac{2\left(1+\varepsilon^2\right)}{\left(1-z^2\right)^2}\,,$$

and

$$e^{-\varepsilon\frac{\pi}{2}} \le |F(z)| \le e^{\varepsilon\frac{\pi}{2}}\,.$$

From these it is easy to see that Equation (7.2) will be satisfied if the right-hand side is replaced by

$$\frac{2+\delta}{\left(1-|z|^2\right)^2}\,,$$

where we can make $\delta > 0$ arbitrarily small if c and ε are each sufficiently small.

For one final example, if we take $p(z) = (1-z^2)^{-1}$ then an extremal map is

$$F(z) = \int_0^z \frac{d\zeta}{\left(1-\zeta^2\right)^2}\,,$$

and similar calculations show that (7.2) will be satisfied for sufficiently small c with equality for z real.

References

[1] Ahlfors, L. Cross-ratios and Schwarzian derivatives in $\mathbb{R}^n$, in *Complex Analysis: Articles Dedicated to Albert Pfluger on the Occasion of his 80th Birthday,* Hersch, J. and Huber, A., Eds., Birkhäuser Verlag, Basel, 1–15, (1989).

[2] Chuaqui, M., Duren, P., and Osgood, B. The Schwarzian derivative for harmonic mappings, *J. Anal. Math.* **91**, 329–351, (2003).

[3] Chuaqui, M., Duren, P., and Osgood, B. Curvature properties of planar harmonic mappings, *Comput. Methods Funct. Theory* **4**, 127–142, (2004).

[4] Chuaqui, M., Duren, P., and Osgood, B. Ellipses, near ellipses, and harmonic Möbius transformations, *Proc. Amer. Math. Soc.* **133**, 2705–2710, (2005).

[5] Chuaqui, M., Duren, P., and Osgood, B. Schwarzian derivatives and uniform local univalence, preprint, (2006).

[6] Chuaqui, M. and Gevirtz, J. Simple curves in $\mathbb{R}^n$ and Ahlfors' Schwarzian derivative, *Proc. Amer. Math. Soc.* **132**, 223–230, (2004).

[7] Chuaqui, M. and Osgood, B. An extension of a theorem of Gehring and Pommerenke, *Israel J. Math.* **91**, 393–407, (1995).

[8] Chuaqui, M. and Osgood, B. Finding complete conformal metrics to extend conformal mappings, *Indiana Univ. Math. J.* **47**, 1273–1291, (1998).

[9] Chuaqui, M. and Osgood, B. General univalence criteria in the disk: Extensions and extremal functions, *Ann. Acad. Sci. Fenn. Math.* **23**, 101–132, (1998).

[10] Dierkes, U., Hildebrandt, S., Küster, A., and Wohlrab, O. *Minimal Surfaces I: Boundary Value Problems,* Springer-Verlag, New York, 1992.

[11] do Carmo, M. *Differential Geometry of Curves and Surfaces,* Prentice-Hall, Englewood Cliffs, NJ, 1976.

[12] Duren, P. *Harmonic Mappings in the Plane,* Cambridge University Press, Cambridge, UK, 2004.

[13] Essén, M. and Keogh, F. The Schwarzian derivative and estimates of functions analytic in the unit disk, *Math. Proc. Camb. Phil. Soc.* **78**, 501–511, (1975).

[14] Gehring, F.W. and Pommerenke, Ch. On the Nehari univalence criterion and quasicircles, *Comment. Math. Helv.* **59**, 226–242, (1984).

[15] Hille, E. Remarks on a paper by Zeev Nehari, *Bull. Amer. Math. Soc.* **55**, 552–553, (1949).

[16] Kraus, W. Über den Zusammenhang einiger Characteristiken eines einfach zusammenhängenden Bereiches mit der Kreissabbildung, *Mitt. Math. Sem. Giessen* **21**, 1–28, (1932).

[17] Nehari, Z. The Schwarzian derivative and schlicht functions, *Bull. Amer. Math. Soc.* **55**, 545–551, (1949).

[18] Nehari, Z. Some criteria of univalence, *Proc. Amer. Math. Soc.* **5**, 700–704, (1954).

[19] Nehari, Z. Univalence criteria depending on the Schwarzian derivative, *Illinois J. Math.* **23**, 345–351, (1979).

[20] O'Neill, B. *Semi-Riemannian Geometry: With Applications to Relativity,* Academic Press, New York, 1983.

[21] Osgood, B. Old and new on the Schwarzian derivative, in *Quasiconformal Mappings in Analysis,* Palka, B., Duren, P., Heinonen, J., and Osgood, B., Eds., Springer-Verlag, New York, 275–308, 1998.

[22] Osgood, B. and Stowe, D. A generalization of Nehari's univalence criterion, *Comm. Math. Helv.* **65**, 234–242, (1990).

[23] Osgood, B. and Stowe, D. The Schwarzian derivative and conformal mapping of Riemannian manifolds, *Duke Math. J.* **67**, 57–97, (1992).

[24] Sheil-Small, T. Constants for planar harmonic mappings, *J. London Math. Soc.* **42**, 237–248, (1990).

Received July 11, 2006

Universidad Católica de Chile
e-mail: mchuaqui@mat.puc.cl

University of Michigan
e-mail: duren@umich.edu

Stanford University
e-mail: osgood@stanford.edu

Communicated by Steven Bell