FUNDAMENTA
MATHEMATICAE
206 (2009)

Higher order Schwarzian derivatives in interval dynamics

by

O. Kozlovski (Coventry) and **D. Sands** (Orsay)

Abstract. We introduce an infinite sequence of higher order Schwarzian derivatives closely related to the theory of monotone matrix functions. We generalize the classical Koebe lemma to maps with positive Schwarzian derivatives up to some order, obtaining control over derivatives of high order. For a large class of multimodal interval maps we show that all inverse branches of first return maps to sufficiently small neighbourhoods of critical values have their higher order Schwarzian derivatives positive up to any given order.

1. Introduction. Many results on interval map dynamics were first proved supposing that the map in question has negative Schwarzian derivative. This is a convexity condition and as such globally constrains the possible shape of the map. Since iterates also have negative Schwarzian derivative, estimates on the distortion of high iterates are often greatly simplified by this assumption. Having negative Schwarzian does not however give good control over derivatives of high order.

The great historical failing of the negative Schwarzian theory was that many interesting and otherwise well-behaved interval maps simply do not have negative Schwarzian derivative everywhere. This flaw was rectified by the discovery that a large class of interval maps are, under some mild hypotheses, real-analytically conjugate to maps with negative Schwarzian derivative everywhere [GS]. An early manifestation of this was the proof [K] that first return maps to small neighbourhoods of critical values have negative Schwarzian derivative.

In this paper we introduce an infinite sequence of higher order Schwarzian derivatives and prove that inverse branches of first return maps to sufficiently small neighbourhoods of critical values have positive Schwarzian derivatives up to some given finite order. (For the classical Schwarzian derivative one

2010 *Mathematics Subject Classification*: 37E05, 53A55, 41A20, 30C20.
Key words and phrases: interval dynamics, Schwarzian derivative, rational approximation, Pick class.

DOI: 10.4064/fm206-1-12

can equally well consider branches having negative Schwarzian derivative or inverse branches having positive Schwarzian derivative: these are equivalent. For higher order Schwarzian derivatives this symmetry disappears, and the natural concept turns out to be positivity.) We also extend the celebrated real Koebe lemma to maps with all Schwarzian derivatives positive up to some order, obtaining control over the distortion of derivatives of high order.

Let d be a positive integer and f a map which is $2d+1$ times differentiable at x. Let $\mathcal{R}$ be the rational map of degree at most d that coincides with f to order $2d$ at x, i.e. for which $\mathcal{R}(x) = f(x)$, $D\mathcal{R}(x) = Df(x)$, …, $D^{2d}\mathcal{R}(x) = D^{2d}f(x)$. Such a rational map may not exist, but if it does exist then it is unique. It is called the dth (*diagonal*) *Padé approximant* [P] to f at x. The *Schwarzian derivative* of f at x of order d is then defined to be

$$S_d(f)(x) \equiv D^{2d+1}(\mathcal{R}^{-1} \circ f)(x).$$

Only a local inverse of $\mathcal{R}$ being needed, this makes sense as long as $Df(x) \neq 0$. It is well-known and easily checked [MS, Lemma 4.4], [T], that this definition gives the classical Schwarzian derivative

$$S_1(f)(x) = \frac{D^3 f(x)}{Df(x)} - \frac{3}{2}\left(\frac{D^2 f(x)}{Df(x)}\right)^2$$

when $d = 1$. Define $S_0(f) \equiv 1$ for convenience.

In order to simplify the exposition, existence of higher order Schwarzian derivatives will be indicated implicitly with the convention that an expression such as $S_d(f)(x) < 0$ is short-hand for: $S_d(f)(x)$ exists and $S_d(f)(x) < 0$.

It is essential to the theory of the classical Schwarzian derivative that iterates of maps with negative Schwarzian derivative also have negative Schwarzian derivative, as follows from the composition formula $S_1(f \circ g) = S_1(f) \circ g\,(Dg)^2 + S_1(g)$. There is also a composition formula for the higher order Schwarzian derivative S_d (Lemma 9), but it contains an extra term coming from the fact that the set of rational maps of degree $d > 1$ is not closed under composition. This term disappears when composing with a Möbius transformation M (i.e. a rational map of degree 1): post-composition has no effect while pre-composition yields $S_d(f \circ M) = S_d(f) \circ M\,(DM)^{2d}$. In the special case of maps with all Schwarzian derivatives of order less than d non-negative, the extra term is non-negative (Proposition 3). This makes estimating the dth order Schwarzian derivative of a long composition of such functions feasible. It also shows that this class of maps is closed under composition. The ultimate origin of this is the fact (Corollary 6) that $S_1(f)(x) > 0, \dots, S_{d-1}(f)(x) > 0$ if and only if the dth Padé approximant to f at x exists, has degree d, and maps the complex upper half-plane into itself (if $Df(x) > 0$) or into the complex lower half-plane (if $Df(x) < 0$).

The following result explains why inverse branches of many well-known one-dimensional maps (logistic maps for example) have Schwarzian derivatives of all orders positive: they map the complex upper half-plane into itself. This is almost but not quite the same as being in the Epstein class. Given an open interval U, let $P_d(U)$ consist of those functions $f : U \to \mathbb{R}$ with $2d+1$ derivatives, $Df \neq 0$ and $S_1(f) \geq 0, \ldots, S_d(f) \geq 0$ everywhere; set $P_\infty(U) = \bigcap_{d=1}^{\infty} P_d(U)$.

PROPOSITION 1. *Let $\phi : U \to \mathbb{R}$ be an increasing C^∞ diffeomorphism onto its image, where U is an open real interval. Then $\phi \in P_\infty(U)$ if and only if ϕ extends to a holomorphic map $\phi : \mathbb{C} \setminus (\mathbb{R} \setminus U) \to \mathbb{C}$ which maps the complex upper half-plane into itself* (*i.e. the extension is in the Pick class*).

It has often been observed that smooth maps become "increasingly holomorphic" when iterated. In the light of the preceding proposition, one way of formalizing this observation is to say that Schwarzian derivatives of ever increasing order become positive for inverse branches under iteration. Our main result shows that this is indeed the case near critical values:

THEOREM 1. *Let $f : I \to I$ be a C^{2d+1} map of a non-trivial compact interval*, *where d is a positive integer*, *and let all critical points of f be non-flat. Then for any critical point c of f which is not in the basin of a periodic attractor there exists a neighbourhood X of c such that if $f^s(x) \in X$ for some $x \in I$ and $s \geq 0$ with $Df^{s+1}(x) \neq 0$, then $S_k(f^{-(s+1)})(f^{s+1}(x)) > 0$ for all $k = 1, \ldots, d$* (*the local inverse of f^{s+1} near x is taken*).

Recall that a critical point c is said to be *non-flat* if f can be decomposed near c as $f = \psi \circ P \circ \phi$ where ϕ (resp. ψ) is a C^{2d+1} diffeomorphism from a neighbourhood of c (resp. 0) onto a neighbourhood of 0 (resp. $f(c)$) and $|P(x)| = |x|^\alpha$ for some $\alpha > 1$ and all small x.

We have so far been unable to prove, when $d > 1$, the global result corresponding to Theorem 1, namely that if all periodic points are hyperbolic repelling then f can be real-analytically conjugated to a map with all Schwarzian derivatives of order d or less positive everywhere for all inverse branches (the conjugacy would depend on d).

The celebrated Koebe lemma for univalent maps $f : U \to V$, where U and V are simply connected domains in $\mathbb{C}$, states that f has bounded distortion on any simply connected domain A which is compactly contained in U: if $x \in A$ and $y \in A$ then $|Df(x)|/|Df(y)|$ is bounded by a constant depending only on the modulus of $U \setminus A$. Higher derivatives are also controlled: after appropriately normalizing the domain U, the ratio $|D^k f(x)|/|Df(x)|$ is again bounded by a constant depending only on k and the modulus of $U \setminus A$.

A real counterpart to the complex Koebe lemma exists for maps with negative classical Schwarzian derivative [MvS, p. 258] but lacks the control of higher order derivatives. Here we show that the real Koebe lemma can be naturally generalized, and control of higher derivatives achieved, in a similar way to the complex Koebe lemma:

THEOREM 2. *Let d be a positive integer, U an open interval and m and n integers such that n is odd and $1 \leq n \leq m \leq 2d$. If $f \in P_d(U)$ then*

$$|D^m f(x)| \leq \frac{m!}{n!}\operatorname{dist}(x, \partial U)^{-(m-n)} |D^n f(x)| \tag{1}$$

for every x in U, where $\operatorname{dist}(x, \partial U)$ *is the distance from x to the boundary of U. The constant in* (1) *is exact and is achieved on a Möbius transformation.*

Note that the inequality in the theorem has the right scaling properties with respect to affine coordinate changes. A similar result [ST, Theorem 1] is known for iterates of general smooth maps, but with additional hypotheses on the sums of lengths of intervals under iteration. Since the class P_d is closed under composition, Theorem 2 applies to arbitrary compositions of maps from P_d without the need for dynamical hypotheses of this kind. This is analogous to the relationship between the classical real Koebe lemma for maps with negative Schwarzian derivative [MvS, p. 258] and that for general smooth maps [MvS, Theorem IV.3.1].

Many additional useful facts about maps in P_d can be deduced from the theory of monotone matrix functions [D]. This is because maps in P_d with positive derivative turn out (Lemma 10) to be exactly the monotone matrix functions of order $d+1$. Consider for example Theorem VII.V from [D]. Dividing the rows and columns of the Pick matrix appropriately turns the matrix elements into cross-ratios; cross-ratios being unaffected by sign changes, the resulting matrix is positive for any function in P_d, not just those that are increasing. The result is a generalization of the well-known cross-ratio contraction property of maps with positive Schwarzian derivative:

PROPOSITION 2. *Let d be a positive integer, U an open interval and take distinct points $\lambda_1, \ldots, \lambda_{d+1}$ in U. If $f \in P_d(U)$ then all eigenvalues of the matrix*

$$\left[\sqrt{\frac{\left(\frac{f(\lambda_i)-f(\lambda_j)}{\lambda_i-\lambda_j}\right)^2}{Df(\lambda_i)Df(\lambda_j)}}\right] \tag{2}$$

(*diagonal elements are equal to* 1) *are non-negative*: *the matrix is positive.*

Some other properties likely to be useful for dynamics can be found in Theorem VII.II, Chapter XIV and Theorem II.I of [D].

Organization. Section 2 sketches the essentials of the theory of Padé approximation as used in this paper and describes some elementary properties of higher order Schwarzian derivatives. Section 3 introduces the Pick algorithm, a degree reduction technique useful for proving results by induction on the order of the Schwarzian derivative. This technique is used in Section 4 to characterize the real rational maps that preserve the complex upper half-plane in terms of their higher order Schwarzian derivatives at a point. At this point the theory is sufficiently developed that obtaining an effective composition formula is trivial—this is done in Section 5. Another application is the proof in Section 6 that the increasing functions in P_d are exactly the monotone matrix functions of order $d+1$. Proposition 1 is thus a restatement of Loewner's Theorem [D]. Section 7 uses the theory of monotone matrix functions to deduce the generalized Koebe lemma from the integral representation for Pick functions. Finally, in Section 8, we prove the main theorem using the a priori bounds of [SV] and the results from [ST] on Epstein class approximation.

2. Rational approximation. This section contains a quick introduction to the classical theory of rational approximation. Most of the results are reformulations of well-known properties.

A *rational map* is a fraction $\mathcal{R} = p/q$ where p and q are polynomials and q is not identically zero. We consider two rational maps $\mathcal{R}_1 = p_1/q_1$ and $\mathcal{R}_2 = p_2/q_2$ to be equal if the polynomials p_1q_2 and p_2q_1 are equal. This means that common polynomial factors in the numerator and the denominator can be cancelled without changing the rational map. When viewing $\mathcal{R}$ as a function we will suppose that p and q are relatively prime, i.e. that $z \mapsto p(z)/q(z)$ has no removable singularities on the Riemann sphere. The *degree* of $\mathcal{R}$, denoted $\deg \mathcal{R}$, is the maximum of the degrees of p and q when p and q are relatively prime.

Two functions f and g are said to *coincide to order* N at some point x if $f(x) = g(x)$, $Df(x) = Dg(x)$, $\ldots$, $D^N f(x) = D^N g(x)$.

LEMMA 1 (Uniqueness). *Let $\mathcal{R}_1$ (resp. $\mathcal{R}_2$) be a rational map of degree at most d_1 (resp. d_2). If $\mathcal{R}_1$ and $\mathcal{R}_2$ are finite at some point x, and coincide to order $d_1 + d_2$ there, then $\mathcal{R}_1 = \mathcal{R}_2$.*

Proof. See [B, Theorem 1.1]. Without loss of generality $x = 0$. Let $\mathcal{R}_i = p_i/q_i$ where p_i, q_i are polynomials of degree at most d_i, $i = 1, 2$. The hypothesis that $\mathcal{R}_1$ and $\mathcal{R}_2$ coincide to order $d_1 + d_2$ is equivalent to $p_1(z)/q_1(z) - p_2(z)/q_2(z) = \mathrm{O}(z^{d_1+d_2+1})$. Multiplying through by the denominators gives $p_1(z)q_2(z) - p_2(z)q_1(z) = \mathrm{O}(z^{d_1+d_2+1})$. The left-hand side is a polynomial of degree at most $d_1 + d_2$ so must in fact be identically zero: $p_1q_2 = p_2q_1$. ∎

Let d be a non-negative integer and f a map which is $2d$ times differentiable at x. Recall that the dth (diagonal) Padé approximant to f at x, denoted $[f]_x^d$, is the rational map of degree at most d that coincides with f to order $2d$ at x, if such a rational map exists. By Lemma 1 there is at most one such rational map.

The classical sufficient condition for the existence of the dth Padé approximant is the non-vanishing of a certain Hankel determinant. The empty determinant is considered equal to 1.

LEMMA 2 (Existence). *Let d be a non-negative integer and x some point. Given numbers $f_0, f_1, \ldots, f_{2d}$ there exists a rational map $\mathcal{R}$ of degree exactly d such that $\mathcal{R}(x) = f_0$, $D\mathcal{R}(x) = f_1$, $\ldots$, $D^{2d}\mathcal{R}(x) = f_{2d}$ if and only if*

$$\det \begin{bmatrix} \frac{f_1}{1!} & \frac{f_2}{2!} & \cdots & \frac{f_d}{d!} \\ \frac{f_2}{2!} & \frac{f_3}{3!} & \cdots & \frac{f_{d+1}}{(d+1)!} \\ \vdots & \vdots & & \vdots \\ \frac{f_d}{d!} & \frac{f_{d+1}}{(d+1)!} & \cdots & \frac{f_{2d-1}}{(2d-1)!} \end{bmatrix} \neq 0. \tag{3}$$

Proof. See [B, Theorem 2.3, points (1) and (5)] or [NS, Chapter 2, Proposition 3.2]. Without loss of generality $x = 0$. Write $F_k = f_k/k!$ and let $f(z) = F_0 + F_1 z + \cdots + F_{2d} z^{2d} + \mathrm{O}(z^{2d+1})$. The existence of an appropriate $\mathcal{R}$, but of degree at most (rather than exactly) d, is equivalent to the existence of a pair (p, q) of polynomials such that

$$\begin{gathered} \operatorname{degree} p \leq d, \qquad \operatorname{degree} q \leq d, \qquad q(0) = 1, \\ F(z)q(z) - p(z) = \mathrm{O}(z^{2d+1}). \end{gathered} \tag{4}$$

Writing $q(z) = 1 + Q_1 z + \cdots + Q_d z^d$, $p(z) = P_0 + P_1 z + \cdots + P_d z^d$ and expressing equation (4) in terms of coefficients yields

$$0 = F_l + \sum_{i=1}^{d} Q_i F_{l-i}, \qquad l = d+1, \ldots, 2d, \tag{5}$$

$$P_l = F_l + \sum_{i=1}^{l} Q_i F_{l-i}, \qquad l = 0, \ldots, d. \tag{6}$$

The key is equation (5), since if it can be solved for $Q_1, \ldots, Q_d$ then the values of $P_0, \ldots, P_d$ are determined by (6). Thus equation (4) has a unique solution if and only if the linear part of (5) is invertible. But this is the case if and only if it has non-zero determinant, which is exactly (3).

The preliminaries now being in place, let us prove the result.

Suppose (3) holds; then (4) has a unique solution (p, q). Uniqueness implies that at least one of p, q has degree d, since if not, the pair

$(p(z)(1+z), q(z)(1+z))$ would be another solution to (4). Uniqueness also implies that p and q are relatively prime since otherwise another solution could be obtained by cancelling common factors (this reasoning uses $q(0) \neq 0$).

Conversely, suppose $\mathcal{R}$ exists with degree exactly d. Then there exists a solution (p, q) of (4) where p and q are relatively prime polynomials, and one of p, q has degree d. Suppose there exists another solution $(\tilde{p}, \tilde{q})$ of (4). Since $\tilde{p}q = p\tilde{q}$ and p and q are relatively prime, there must exist some non-zero polynomial r such that $\tilde{p} = rp$, $\tilde{q} = rq$. Now $\tilde{p}$ and $\tilde{q}$ have degree at most d while one of p, q has degree d, so r must have degree zero, i.e. be a constant. From $q(0) = 1 = \tilde{q}(0)$ it follows that the constant r is in fact 1, i.e. $\tilde{p} = p$ and $\tilde{q} = q$. In other words, (4) has a unique solution, which means that (3) holds. ∎

Write

$$M_d(x, f) \equiv \begin{bmatrix} \frac{Df(x)}{1!} & \cdots & \frac{D^d f(x)}{d!} \\ \vdots & & \vdots \\ \frac{D^d f(x)}{d!} & \cdots & \frac{D^{2d-1} f(x)}{(2d-1)!} \end{bmatrix}. \tag{7}$$

Corollary 1. *Let d be a non-negative integer, U an open interval and $f : U \to \mathbb{R}$ a function with $2d + 1$ derivatives. Then* $\det M_{d+1}(x, f) = 0$ *for all $x \in U$ if and only if f is a rational map of degree at most d.*

Proof. Suppose f is a rational map of degree at most d. If $\det M_{d+1}(x, f) \neq 0$ then by Lemma 2 there is a rational map $\mathcal{R}$ of degree $d + 1$ which coincides with f to order $2d + 2$ at x. Then Lemma 1 shows that $f = \mathcal{R}$, which is impossible because their degrees differ. Thus $\det M_{d+1}(x, f) = 0$ for all $x \in U$. For the converse, the difficulty is that the differential equation to be solved is singular at points where $\det M_d(x, f) = 0$.

Claim. *Suppose* $\det M_d(x, f) \neq 0$, $\det M_{d+1}(x, f) = 0$ *for all x in some open interval $V \subseteq U$. Then f is a rational map of degree d on V.*

Proof of Claim. Observe that $\det M_{d+1}(x, f) = 0$ is an ordinary differential equation for f with highest term $D^{2d+1}(f)(x)$. The coefficient of this term is $\det M_d(x, f)$, which is non-zero by hypothesis, so the differential equation is non-singular: there is local existence and uniqueness. Let $\mathcal{R}$ be the Padé approximant to f of order d at some point p of V ($\mathcal{R}$ exists and has degree d by Lemma 2). Note that $\det M_{d+1}(x, \mathcal{R}) = 0$ for all $x \in V$, i.e. $\mathcal{R}$ is a solution of the differential equation (this was proved above). By definition $\mathcal{R}$ coincides with f to order $2d$ at p, so f and $\mathcal{R}$ have the same initial values at p and thus coincide throughout V. This proves the Claim.

So suppose $\det M_{d+1}(x, f) = 0$ for all $x \in U$. First consider the case when $M_d(x, f) \neq 0$ for some point $x \in U$ (this is always the case if $d = 0$).

Then $M_d(\cdot, f) \neq 0$ at every point of U. Indeed, let V be the maximal open interval around x on which $M_d(\cdot, f) \neq 0$. By the Claim, f coincides with a rational map $\mathcal{R}$ of degree d on V. If $V \neq U$ then there is some boundary point p of V in U. By continuity, $D^k(\mathcal{R})(p) = D^k(f)(p)$ for $0 \leq k \leq 2d$, so $\det M_d(p, f) \neq 0$ by Lemma 2. This implies that $p \in V$, a contradiction since V is open. Thus $V = U$ and f is a rational map of degree d.

Now suppose that $\det M_d(x, f) = 0$ for all $x \in U$. This reduces d by 1 in the hypotheses. By induction f is a rational map of degree at most $d-1$. ∎

Definition 1. We say that f is *normal of order* d at x if the dth Padé approximant to f at x exists and has degree exactly d.

Corollary 2. *Let f have $2d$ derivatives at x. Then f is normal of order d at x if and only if* $\det M_d(x, f) \neq 0$.

If f is not normal of order d then the dth Padé approximant may nonetheless exist, but if so it is simply equal to the $(d-1)$st Padé approximation:

Corollary 3. *Let f have $2d+2$ derivatives at x. Suppose the $(d+1)$st Padé approximant $[f]_x^{d+1}$ exists. If f is not normal of order $d+1$ at x then the dth Padé approximant $[f]_x^d$ exists and $[f]_x^{d+1} = [f]_x^d$. If f is normal of order $d+1$ at x then either f is normal of order d at x, or the dth Padé approximant to f at x does not exist.*

Proof. See also [B, Theorem 2.3]. If f is not normal of order $d+1$ then $[f]_x^{d+1}$ has degree at most d. Since it satisfies the conditions to be the dth Padé approximant, by uniqueness it is $[f]_x^d$. Now suppose f is normal of order $d+1$ and the dth Padé approximant $[f]_x^d$ exists. If f is not normal of order d then $[f]_x^d$ has degree at most $d-1$. Since $[f]_x^d$ and $[f]_x^{d+1}$ coincide to order $2d$ at x, as both coincide with f to at least that order, they are equal by Lemma 1. This contradicts $[f]_x^{d+1}$ having degree $d+1$. ∎

Another way of viewing this result is as follows: Suppose f is normal of order d but not of order $d+1$. Let $N \in \{d+1, d+2, \ldots\}$ be minimal such that the Nth Padé approximant does not exist, or set $N = \infty$ if Padé approximants exist of all orders. Then the approximants of orders $d < k < N$ are all equal to $[f]_x^d$.

Recall that the Schwarzian derivative of f at x of order d is defined to be $S_d(f)(x) = D^{2d+1}(([f]_x^d)^{-1} \circ f)(x)$. An alternative definition is $S_d(f)(x) = D^{2d+1}(f - [f]_x^d)(x)/Df(x)$. The equivalence of the two expressions is readily derived by induction from the fact that f and $[f]_x^d$ coincide to order $2d$ at x.

Lemma 3 (Schwarzian formula). *Let d be a non-negative integer and x some point. If f has $2d+1$ derivatives at x, $Df(x) \neq 0$, and f is normal of* *order d at x, then*

$$S_d(f)(x) = (2d+1)!\,\frac{\det M_{d+1}(x,f)}{Df(x)\det M_d(x,f)}. \tag{8}$$

Proof. Let $\mathcal{R} = [f]_x^d$ be the dth Padé approximant to f at x. Recall the alternative definition $S_d(f)(x) = (D^{2d+1}f(x) - D^{2d+1}\mathcal{R}(x))/Df(x)$ of the Schwarzian derivative of order d. Write

$$A = \begin{bmatrix} F_1 & \cdots & F_d \\ \vdots & & \vdots \\ F_d & \cdots & F_{2d-1} \end{bmatrix}, \tag{9}$$

$$C = [F_{d+1} \quad \cdots \quad F_{2d}] \tag{10}$$

and $D = F_{2d+1}$, where $F_k = D^k f(x)/k!$. Let $\mathcal{R} = p/q$ where $p(z) = P_0 + P_1 z + \cdots + P_d z^d$ and $q(z) = 1 + Q_1 z + \cdots + Q_d z^d$ are polynomials of degree at most d. Equation (5) can be rewritten as $[Q_d, \ldots, Q_1]^T = -A^{-1}C^T$. Since $D^{2d+1}\mathcal{R}(x)/(2d+1)! = -C[Q_d, \ldots, Q_1]^T$, it follows that $(D^{2d+1}f(x) - D^{2d+1}\mathcal{R}(x))/(2d+1)! = D - CA^{-1}C^T$. Applying the well-known formula

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det A \det(D - CA^{-1}B) \tag{11}$$

for the determinant of a conformally partitioned block matrix with $B = C^T$ immediately yields

$$D - CA^{-1}C^T = \frac{\det M_{d+1}(x,f)}{\det M_d(x,f)}. \tag{12}$$

Thus

$$S_d(f)(x) = \frac{D^{2d+1}f(x) - D^{2d+1}\mathcal{R}(x)}{Df(x)} = (2d+1)!\,\frac{\det M_{d+1}(x,f)}{Df(x)\det M_d(x,f)}. \tag{13}$$ ∎

LEMMA 4. *Let d be a non-negative integer, U an open interval and $f : U \to \mathbb{R}$ a function with $2d+1$ derivatives. Suppose $Df(x) \neq 0$ for all $x \in U$. Then $S_d(f)$ is identically zero on U if and only if f is a rational map of degree at most d.*

Proof. If f is a rational map of degree at most d, then f is the dth Padé approximant to itself at any point, so $S_d(f) = 0$ by definition. So suppose $S_d(f)$ is identically zero on U. If f is not normal of order d at any point of U (this does not happen if $d = 0$) then $\det M_d(x,f) = 0$ for all $x \in U$, so f is a rational map of degree at most $d-1$ by Corollary 1. So suppose f is normal of order d at some point x of U. Then, as in the proof of Corollary 1, f is normal of order d at every point of U. Indeed, let V be the maximal open interval around x on which f is normal of order d. Then formula (8) is valid on V, so from $S_d(f) \equiv 0$ it follows that $\det M_{d+1}(\cdot, f) = 0$ on V. It was shown in the proof of Corollary 1 that f then coincides on V with a rational map $\mathcal{R}$ of degree exactly d. If $V \neq U$ then there is some boundary point p of V in U. By continuity, f and $\mathcal{R}$ coincide to order $2d$ at p, so $\mathcal{R}$ is the dth Padé approximant to f at p. Thus f is normal of order d at p by definition, i.e. $p \in V$. This contradicts V being open. Hence $V = U$ and f is a rational map of degree d. ■

3. The Pick algorithm. Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ be the open unit disk in the complex plane. The *Schur class* consists of all holomorphic functions $\mathbb{D} \to \overline{\mathbb{D}}$. Given a holomorphic map $F : \mathbb{D} \to \mathbb{D}$, the Schur algorithm generates a new holomorphic map $\tilde{F} : \mathbb{D} \to \overline{\mathbb{D}}$ as follows. Let M_F be the Möbius transformation $z \mapsto (z - F(0))/(1 - \overline{F(0)}z)$. This preserves $\mathbb{D}$ and maps $F(0)$ to 0. The function $z \mapsto M_F(F(z))/z$ has a removable singularity at $z = 0$, so extends to a holomorphic function $\tilde{F} : \mathbb{D} \to \mathbb{C}$. The Schwarz lemma shows that in fact $\tilde{F} : \mathbb{D} \to \overline{\mathbb{D}}$. If $\tilde{F}$ is not constant then $\tilde{F} : \mathbb{D} \to \mathbb{D}$. Applying the algorithm iteratively results in a finite or infinite sequence of Schur maps that terminates with a constant function if finite.

The Schur algorithm uses 0 as a distinguished point of $\mathbb{D}$ and a particular choice of Möbius transformation taking $F(0)$ to 0. Other choices produce different sequences of maps. Moving the distinguished point towards the boundary of $\mathbb{D}$, normalizing with Möbius transformations, and passing to the limit results in a version of the Schur algorithm for which the distinguished point lies on the unit circle. This is the Pick algorithm studied in this section.

Rather than work with $\mathbb{D}$, $\overline{\mathbb{D}}$ and the unit circle, it is more convenient to use the conformally equivalent complex upper half-plane $\mathbb{H} = \{z \in \mathbb{C} : \Im(z) > 0\}$, its closure in the Riemann sphere $\overline{\mathbb{H}} = \{z \in \mathbb{C} : \Im(z) \geq 0\} \cup \{\infty\}$, and the extended real line $\mathbb{R} \cup \{\infty\}$. Section 4 makes use of the Pick algorithm in the complex plane. Here we consider real-valued functions defined on a real neighbourhood of a point $x \in \mathbb{R}$ since this suffices for our applications.

DEFINITION 2. Let f be twice differentiable at x with $Df(x) \neq 0$. The *Pick algorithm* based at x transforms f into

$$\mathcal{P}_x(f) : z \mapsto \begin{cases} \dfrac{1 - Df(x)\,\frac{z-x}{f(z)-f(x)}}{z - x}, & z \neq x, \\[2ex] \dfrac{D^2 f(x)}{2\,Df(x)}, & z = x. \end{cases} \tag{14}$$

Note that $\mathcal{P}_x(f)$ is continuous at x (in general, two derivatives are lost at x). If $f(x)$ and $Df(x)$ are known, then f can be recovered from $\tilde{f} = \mathcal{P}_x(f)$:

Definition 3. Let $\tilde{f}$ be continuous at x, and take some $A \in \mathbb{R}$ and $\mu \in \mathbb{R} \setminus \{0\}$. The *inverse Pick algorithm* based at x transforms $\tilde{f}$ into

$$f : z \mapsto A + \frac{\mu(z-x)}{1-(z-x)\tilde{f}(z)}. \tag{15}$$

This is indeed an inverse: $\mathcal{P}_x(f) = \tilde{f}$. Note that f is twice differentiable at x, $f(x) = A$ and $Df(x) = \mu \neq 0$ (in general, two derivatives are gained at x).

Clearly, f is a rational map if and only if $\mathcal{P}_x(f)$ is. If they are rational, it is straightforward to show that $\deg f = 1 + \deg \mathcal{P}_x(f)$. This uses the standing assumptions that f is real (which implies that $f(x)$ is finite) and $Df(x) \neq 0$.

It can be helpful to think of the Pick algorithm in terms of continued fractions. Applying the inverse Pick algorithm d times to $\tilde{f}$ results in the Jacobi-type continued fraction

$$f : z \mapsto A_0 + \cfrac{\mu_0(z-x)}{1-(z-x)A_1 - \cfrac{\mu_1(z-x)^2}{1-(z-x)A_2 - \cfrac{\mu_2(z-x)^2}{\ddots - \cfrac{\mu_{d-1}(z-x)^2}{1-(z-x)\tilde{f}(z)}}}} \tag{16}$$

$$A_0, \dots, A_{d-1} \in \mathbb{R}, \qquad \mu_0, \dots, \mu_{d-1} \in \mathbb{R} \setminus \{0\}.$$

Since we only use continued fractions to illustrate results rather than prove them, we have felt free to state their properties without justification. Observe that f is $2d$ times differentiable at x and $\mathcal{P}_x^d(f) = \tilde{f}$. More: f is normal of orders $1, \dots, d$ at x. Conversely, if f is $2d$ times differentiable at x and f is normal of orders $1, \dots, d$ at x then f can be written in the form (16) with $\tilde{f}$ continuous at x. The kth convergent of equation (16) (obtained by setting $\mu_k = 0$) is exactly the kth Padé approximant to f at x ($0 \le k < d$). These properties are the essence of:

Lemma 5. *Let d be a positive integer and f a map which is $2d$ times differentiable at x with $Df(x) \neq 0$. The dth Padé approximant $[f]_x^d$ to f at x exists if and only if the $(d-1)$st Padé approximant $[\mathcal{P}_x(f)]_x^{d-1}$ to $\mathcal{P}_x(f)$ exists, and then $\mathcal{P}_x([f]_x^d) = [\mathcal{P}_x(f)]_x^{d-1}$.*

Proof. Write $\tilde{f}$ for $\mathcal{P}_x(f)$. The case $d = 1$ is immediate: the 0th Padé approximant always exists and the first Padé approximant to f exists because $Df(x) \neq 0$. The formula connecting the two is trivial. So suppose $d > 1$ and that $[\tilde{f}]_x^{d-1}$ exists—denote it by $\mathcal{T}$. Define $\mathcal{R}$ via the inverse Pick algorithm: $\mathcal{R}(z) = f(x) + Df(x)(z-x)/(1-(z-x)\mathcal{T}(z))$. Note that $\deg \mathcal{R} = 1 + \deg \mathcal{T}$.

Then

$$f(z)-\mathcal{R}(z)=\frac{1}{Df(x)}\,\frac{f(z)-f(x)}{z-x}\,\frac{\mathcal{R}(z)-\mathcal{R}(x)}{z-x}\,(z-x)^2(\tilde f(z)-\mathcal{T}(z)). \tag{17}$$

By definition $\tilde f$ and $\mathcal{T}$ coincide to order $2(d-1)$, i.e. $\tilde f(z)-\mathcal{T}(z)=\mathrm{o}((z-x)^{2(d-1)})$. Then $f(z)-\mathcal{R}(z)=\mathrm{o}((z-x)^{2d})$ by (17), which means that f and $\mathcal{R}$ coincide to order $2d$. Thus $\mathcal{R}$ is the dth Padé approximant to f at x, as desired. The case when it is the dth Padé approximant to f at x that is initially known to exist is left to the interested reader. ∎

Corollary 4. *Let d be a positive integer and f a map which is $2d+1$ times differentiable at x with $Df(x)\neq 0$. Then*

$$S_d(f)(x)=2d(2d+1)D\mathcal{P}_x(f)(x)S_{d-1}(\mathcal{P}_x(f))(x). \tag{18}$$

Included in this is that $S_d(f)$ exists if and only if $S_{d-1}(\mathcal{P}_x(f))$ exists.

Proof. The statement about existence is immediate from Lemma 3. For the formula, revisit the proof of the lemma. Writing $\tilde f(z)-\mathcal{T}(z)=\alpha(z-x)^{2d-1}+\mathrm{o}((z-x)^{2d-1})$, we observe that $\alpha=D\tilde f(x)S_{d-1}(\tilde f)(x)/(2d-1)!$. This observation is precisely the alternative definition of the higher Schwarzian derivative from Section 2. Likewise, $f(z)-\mathcal{R}(z)=Df(x)\alpha(z-x)^{2d+1}+\mathrm{o}((z-x)^{2d+1})$—which follows from (17)—means $S_d(f)(x)=\alpha(2d+1)!$. This is the same as (18). ∎

In terms of the continued fraction representation (16), this says that

$$\mu_k=\frac{S_k(f)(x)}{2k(2k+1)S_{k-1}(f)(x)}\qquad\text{for } 1\le k<d.$$

Since $A_1=\frac{1}{2}D^2f(x)/Df(x)$, the A_k can presumably be expressed in terms of higher order non-linearities.

4. Rational Pick maps. In this section we use the Pick algorithm to characterize the real rational maps in the Pick class as those with their Schwarzian derivatives of all orders non-negative. Remarkably, if they are non-negative at a single point then they are non-negative everywhere.

A rational map $\mathcal{R}$ is *real* if it can be written as a ratio of polynomials with only real coefficients. This is equivalent to $\mathcal{R}$ being real- or infinite-valued everywhere on the real line. By uniqueness, Padé approximants to real-valued maps (the only kind of Padé approximant considered in this paper) are real.

Definition 4. The *Pick class* consists of all holomorphic functions $\mathbb{H}\to\overline{\mathbb{H}}$.

Non-constant members of the Pick class map the complex upper half-plane $\mathbb{H}$ into itself, as follows from the open mapping theorem. The one-to-one correspondence between the Schur and Pick classes can be used to transform properties of the Schur class, such as the characterization of Schur rational maps as finite Blaschke products multiplied by numbers in $\overline{\mathbb{D}}$, into statements about the Pick class. But for our purposes it is simpler to work directly with the Pick class.

LEMMA 6. *If a rational map $\mathcal{R}$ is in the Pick class and $\mathcal{R}$ is real-valued (hence finite) at some point $x \in \mathbb{R}$ then either $\mathcal{R}$ is a constant or $D\mathcal{R}(x) > 0$.*

Proof. If $D\mathcal{R}(x) < 0$ then all points $x + \varepsilon i \in \mathbb{H}$ with $\varepsilon > 0$ sufficiently small would be mapped into the lower half-plane. If $D\mathcal{R}(x) = 0$ and $\mathcal{R}$ is not a constant, then $\mathcal{R}(z) = \mathcal{R}(x) + \alpha(z-x)^k + \mathrm{O}((z-x)^{k+1})$ with $k > 1$ and $\alpha \neq 0$, so again some points in $\mathbb{H}$ would be mapped into the lower half-plane. ∎

As the following lemma shows, every real rational Pick map of positive degree can be generated via the inverse Pick algorithm from a real rational Pick map of degree one smaller (μ should be taken positive in the inverse algorithm in order to generate a map with positive derivative at x).

LEMMA 7 (Degree reduction). *Let $\mathcal{R}$ be a real rational map and $x \in \mathbb{R}$ some point for which $\mathcal{R}(x)$ is finite and $D\mathcal{R}(x) > 0$. Then $\mathcal{R}$ is in the Pick class if and only if $\mathcal{P}_x(\mathcal{R})$ is in the Pick class.*

Proof. Let $\mathcal{T} = \mathcal{P}_x(\mathcal{R})$ and recall the relationship

$$\mathcal{R}(z) = \mathcal{R}(x) + \frac{D\mathcal{R}(x)(z-x)}{1-(z-x)\mathcal{T}(z)}. \tag{19}$$

First suppose that $\mathcal{T}$ is in the Pick class and take an arbitrary point $z_0 \in \mathbb{H}$. In order to see that $\mathcal{R}(z_0) \in \mathbb{H}$, let $B = \mathcal{T}(z_0) \in \overline{\mathbb{H}}$ and consider the Möbius transformation

$$M_B : z \mapsto \mathcal{R}(x) + \frac{D\mathcal{R}(x)(z-x)}{1-(z-x)B}. \tag{20}$$

Since $z \mapsto -1/z$ maps $\mathbb{H}$ to $\mathbb{H}$, so does $z \mapsto B - 1/z$ because $\Im(B) \geq 0$. Composing with $z \mapsto -D\mathcal{R}(x)/z$ shows that $z \mapsto z\,D\mathcal{R}(x)/(1-Bz)$ also maps $\mathbb{H}$ into itself since $D\mathcal{R}(x) > 0$. Thus M_B maps $\mathbb{H}$ into itself because x and $\mathcal{R}(x)$ are real. In particular, $\mathcal{R}(z_0) = M_B(z_0) \in \mathbb{H}$. This shows that $\mathcal{R}$ is in the Pick class.

Now suppose that $\mathcal{R}$ is in the Pick class. The Poincaré distance $d(z,w)$ between points z, w of $\mathbb{H}$ is given by

$$d(z,w) = \log \frac{|z-\overline{w}| + |z-w|}{|z-\overline{w}| - |z-w|}. \tag{21}$$

Since $\mathcal{R}$ maps $\mathbb{H}$ holomorphically into itself, it does not expand the Poincaré distance: $d(\mathcal{R}(z),\mathcal{R}(w)) \le d(z,w)$, which is equivalent to

$$\frac{|\mathcal{R}(z)-\overline{\mathcal{R}(w)}|+|\mathcal{R}(z)-\mathcal{R}(w)|}{|z-\overline{w}|+|z-w|} \le \frac{|\mathcal{R}(z)-\overline{\mathcal{R}(w)}|-|\mathcal{R}(z)-\mathcal{R}(w)|}{|z-\overline{w}|-|z-w|}. \tag{22}$$

Writing $w = x+\varepsilon i$ and passing to the limit $\varepsilon \downarrow 0$ in (22) yields

$$\frac{|\mathcal{R}(z)-\mathcal{R}(x)|}{|z-x|} \le D\mathcal{R}(x)\,\frac{\Im(\mathcal{R}(z))}{\Im(z)}\,\frac{|z-x|}{|\mathcal{R}(z)-\mathcal{R}(x)|}, \tag{23}$$

where we have used the equality $|u+v|-|u-v| = 4\Re(u\overline{v})/(|u+v|+|u-v|)$ to evaluate the right-hand side of (22) (with $u = z-x$, $v = \varepsilon i$ in the denominator, and $u = \mathcal{R}(z) - \Re(\mathcal{R}(w))$, $v = \Im(\mathcal{R}(w))i$ in the numerator). By the identity $\Im(\alpha)/|\alpha|^2 = -\Im(1/\alpha)$ used twice, once with $\alpha = z-x$ and once with $\alpha = \mathcal{R}(z)-\mathcal{R}(z)$, this rearranges to

$$0 \le \Im\left(\frac{1}{z-x} - \frac{D\mathcal{R}(x)}{\mathcal{R}(z)-\mathcal{R}(x)}\right), \tag{24}$$

which is exactly $\Im(\mathcal{T}(z)) \ge 0$ since

$$\mathcal{T}(z) = \frac{1-\frac{D\mathcal{R}(x)(z-x)}{\mathcal{R}(z)-\mathcal{R}(x)}}{z-x}. \tag{25}$$ ∎

It is now easy to understand why a real rational map in the Pick class has non-negative Schwarzian derivatives of all orders: applying the Pick algorithm repeatedly gives a sequence of real rational Pick maps of decreasing degree, finishing with a constant. Except for the constant, these all have positive derivative (Lemma 6). But the higher order Schwarzians of the original map are just products of these derivatives, up to a positive constant (Corollary 4).

Lemma 8 (Characterization). *Let $\mathcal{R}$ be a real rational map of degree $d \ge 1$, and $x \in \mathbb{R}$ some point at which $\mathcal{R}$ is finite. If $\mathcal{R}$ is in the Pick class then $D\mathcal{R}(x) > 0$ and $S_k(\mathcal{R})(x) > 0$ for $1 \le k < d$. The existence of the Schwarzian derivatives is part of the conclusion. Conversely, if $D\mathcal{R}(x) > 0$ and $S_k(\mathcal{R})(x) \ge 0$ for $1 \le k < d$ then $\mathcal{R}$ is in the Pick class. The existence of the Schwarzian derivatives is part of the hypotheses.*

Proof. We proceed by induction on the degree. The case $d = 1$ is easily checked, so take $d > 1$. If $\mathcal{R}$ is in the Pick class then $D\mathcal{R}(x) > 0$ (Lemma 6) and $\mathcal{T} \equiv \mathcal{P}_x(\mathcal{R})$ is in the Pick class (Lemma 7). Because $\deg \mathcal{T} = d-1 \ge 1$, it follows by induction that $D\mathcal{T}(x) > 0$ and $S_k(\mathcal{T})(x) > 0$ for $1 \le k < d-1$. Corollary 4 immediately gives $S_k(\mathcal{R})(x) > 0$ for $1 \le k < d$.

Conversely, if $D\mathcal{R}(x) > 0$ and $S_k(\mathcal{R})(x) \ge 0$ for $1 \le k < d$ then, by Corollary 4, $S_k(\mathcal{T})(x) \ge 0$ for $1 \le k < d-1$. Note that $S_1(\mathcal{R})(x) \ne 0$ (otherwise $\mathcal{R}$ would not be normal of order 2; by hypothesis the Padé approximants to $\mathcal{R}$ at x of order $1,\dots,d$ exist, so Corollary 3 would then imply that $\mathcal{R}$ equals $[\mathcal{R}]_x^1$, which has degree 1, a contradiction with $d > 1$), so $D\mathcal{T}(x) > 0$ by Corollary 4. Thus by induction $\mathcal{T}$ is in the Pick class, and therefore also $\mathcal{R}$ by Lemma 7. ∎

COROLLARY 5. *Let $\mathcal{R}$ be a real rational Pick map of degree d which is finite at x. Then $\mathcal{R}$ is normal of orders $0, \ldots, d$ at x.*

Proof. If $\mathcal{R}$ is constant then there is nothing to prove. Otherwise $D\mathcal{R}(x) \neq 0$ by Lemma 6, which means that $\mathcal{R}$ is normal of order 1 at x. Since $S_1(\mathcal{R})(x) \neq 0$ by Lemma 8, it follows from Lemma 3 that $\det M_2(x, \mathcal{R}) \neq 0$, which shows that $\mathcal{R}$ is normal of order 2 at x (Corollary 2). Repeat for higher orders using the non-zero Schwarzian derivatives ensured by Lemma 8. ∎

Thus every real rational Pick map of degree d which is finite at a point $x \in \mathbb{R}$ can be written in the form

$$\mathcal{R}(z) = A_0 + \cfrac{\mu_0(z-x)}{1-(z-x)A_1 - \cfrac{\mu_1(z-x)^2}{1-(z-x)A_2 - \cfrac{\mu_2(z-x)^2}{\ddots \over 1-(z-x)A_d}}} \tag{26}$$

where $A_0, \ldots, A_d \in \mathbb{R}$ and $\mu_0, \ldots, \mu_{d-1}$ are strictly positive. Conversely, every function of this form defines a real rational Pick map of degree d which is finite at x.

COROLLARY 6. *Let d be a positive integer and x some point. Suppose f has $2d$ derivatives at x and $Df(x) > 0$. Then $S_1(f)(x) > 0, \ldots, S_{d-1}(f)(x) > 0$ if and only if the dth Padé approximant to f at x exists, has degree d, and is in the Pick class.*

Proof. If $S_1(f)(x) > 0, \ldots, S_{d-1}(f)(x) > 0$ then, like in the proof of Corollary 5, f is normal of orders $0, \ldots, d$ by induction on the order. In particular, $[f]_x^d$ exists and has degree d. Now suppose that $\mathcal{R} \equiv [f]_x^d$ exists and has degree d. Since f and $\mathcal{R}$ coincide to order $2d$, it is immediate that $S_k(\mathcal{R})(x) = S_k(f)(x)$ for $1 \leq k < d$. Thus $S_1(f)(x) > 0, \ldots, S_{d-1}(f)(x) > 0$ if and only if $S_1(\mathcal{R})(x) > 0, \ldots, S_{d-1}(\mathcal{R})(x) > 0$, and this, according to Lemma 8, if and only if $\mathcal{R}$ is in the Pick class. ∎

5. Composition formula. The composition formula for the classical Schwarzian derivative, $S_1(g \circ f) = S_1(g) \circ f\ (Df)^2 + S_1(f)$, implies that the set of maps for which S_1 is identically zero is closed under composition. This set is precisely the group of Möbius transformations—the composition formula implicitly contains the group structure of these maps. On the other hand, the set of rational maps of degree at most d is not closed under composition, yet these are the maps for which S_d is identically zero (Lemma 4). Inevitably the composition formula for S_d contains additional terms reflecting the lack of group structure:

LEMMA 9. *Let d be a positive integer, f (resp. g) a function which is $2d+1$ times differentiable at x (resp. $f(x)$). Suppose $Df(x) \neq 0$, $Dg(f(x)) \neq 0$ and $S_d(f)(x)$, $S_d(g)(f(x))$ and $S_d(g \circ f)(x)$ exist. Then*

$$S_d(g\circ f)(x) = S_d(g)(f(x))(Df(x))^{2d} + S_d(f)(x) + S_d([g]^d_{f(x)} \circ [f]^d_x)(x). \tag{27}$$

Note that if either $[g]^d_{f(x)}$ or $[f]^d_x$ is a Möbius transformation, then $S_d([g]^d_{f(x)} \circ [f]^d_x)(x) = 0$ because $[g]^d_{f(x)} \circ [f]^d_x$ has degree at most d.

Proof of Lemma 9. The result is almost immediate from the definitions. Indeed, set $\mathcal{F} \equiv [f]^d_x$ and $\mathcal{G} \equiv [g]^d_{f(x)}$; define $\Delta_f = \mathcal{F}^{-1} \circ f$, $\Delta_g = \mathcal{G}^{-1} \circ g$ and $\Delta_\circ = ([\mathcal{G} \circ \mathcal{F}]^d_x)^{-1} \circ \mathcal{G} \circ \mathcal{F}$. By definition

$$\begin{aligned}
\Delta_f(z) &= z + \frac{S_d(f)(x)}{(2d+1)!}(z-x)^{2d+1} + \mathrm{o}((z-x)^{2d+1}),\\
\Delta_g(z) &= z + \frac{S_d(g)(f(x))}{(2d+1)!}(z-f(x))^{2d+1} + \mathrm{o}((z-f(x))^{2d+1}),\\
\Delta_\circ(z) &= z + \frac{S_d(\mathcal{G}\circ\mathcal{F})(x)}{(2d+1)!}(z-x)^{2d+1} + \mathrm{o}((z-x)^{2d+1}).
\end{aligned} \tag{28}$$

Let $\widehat{\Delta_g} = \mathcal{F}^{-1} \circ \Delta_g \circ \mathcal{F}$. Manipulating the Δ_g series and using $D\mathcal{F}(x) = Df(x)$, we get

$$\widehat{\Delta_g}(z) = z + \frac{S_d(g)(f(x))}{(2d+1)!} Df(x)^{2d}(z-x)^{2d+1} + \mathrm{o}((z-x)^{2d+1}). \tag{29}$$

Observe that $g\circ f = \mathcal{G}\circ\Delta_g\circ\mathcal{F}\circ\Delta_f = \mathcal{G}\circ\mathcal{F}\circ\widehat{\Delta_g}\circ\Delta_f = [\mathcal{G}\circ\mathcal{F}]^d_x\circ\Delta_\circ\circ\widehat{\Delta_g}\circ\Delta_f$. Composing the series for last three terms gives

$$\begin{aligned}
&\Delta_\circ \circ \widehat{\Delta_g} \circ \Delta_f(z)\\
&= z + \big(S_d(\mathcal{G}\circ\mathcal{F})(x) + S_d(g)(f(x))Df(x)^{2d} + S_d(f)(x)\big)\frac{(z-x)^{2d+1}}{(2d+1)!}\\
&\qquad + \mathrm{o}((z-x)^{2d+1}).
\end{aligned}$$

Thus $[g\circ f]^d_x = [\mathcal{G}\circ\mathcal{F}]^d_x$ and equation (27) follows. ∎

Nonetheless, for any d there is a general composition inequality for maps with non-negative Schwarzian derivatives of lower order. The reason for this is that Padé approximants to such maps correspond to (non-constant) members of the Pick class, and the set of such Pick maps is closed under composition.

PROPOSITION 3. *Let d be a positive integer and f (resp. g) a function which is $2d+1$ times differentiable at x (resp. $f(x)$). Suppose $Df(x) \neq 0$, $Dg(f(x)) \neq 0$, $S_d(f)(x)$ and $S_d(g)(f(x))$ exist, and $S_k(f)(x) \geq 0$, $S_k(g)(f(x)) \geq 0$ for every $1 \leq k < d$. Then $S_d(g \circ f)(x)$ exists and*

$$S_d(g \circ f)(x) \geq S_d(g)(f(x))(Df(x))^{2d} + S_d(f)(x). \tag{30}$$

Proof. Without loss of generality, $Df(x) > 0$ and $Dg(f(x)) > 0$ (otherwise pre- and/or post-compose with $z \mapsto -z$ to arrange this—Schwarzian derivatives do not change). The dth Padé approximant $\mathcal{F} \equiv [f]_x^d$ to f at x is in the Pick class by Lemma 8 because $S_k(\mathcal{F})(x) = S_k(f)(x) \geq 0$ for $1 \leq k < d$. Likewise, $\mathcal{G} \equiv [g]_{f(x)}^d$ is in the Pick class. In order words, $\mathcal{F}$ and $\mathcal{G}$ map the complex upper half-plane $\mathbb{H}$ into itself. Obviously, their composition does too, i.e. $\mathcal{G} \circ \mathcal{F}$ is in the Pick class. Thus $[\mathcal{G} \circ \mathcal{F}]_x^d$ exists by Lemma 5—this is equivalent to the existence of $S_d(g \circ f)$. Finally, $S_d(\mathcal{G} \circ \mathcal{F})(x) \geq 0$ by Lemma 8. ■

6. Monotone matrix functions. In this section we introduce the class of monotone matrix functions and relate them to maps with non-negative higher order Schwarzian derivatives. A complete description of this class can be found in [D].

Recall how to take the image of a real symmetric matrix A by a function f: if A is diagonal, $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, then $f(A) = \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_n))$; otherwise diagonalize A via some linear coordinate change, take the image of the diagonal matrix, and undiagonalize by applying the inverse coordinate change. This is well-defined if the spectrum of A belongs to the domain of f.

Recall the ordering on the real symmetric n-by-n matrices: $A \leq B$ if and only if $B - A$ is a positive matrix, meaning $v^T(B-A)v \geq 0$ for every n-by-1 vector v.

DEFINITION 5. Let n be a positive integer, U an open interval of the real line, and $f : U \to \mathbb{R}$ a function. Call f *matrix monotone of order* n if, for any real symmetric n-by-n matrices A and B with spectrum contained in U, $A \leq B$ implies $f(A) \leq f(B)$.

LEMMA 10. *Let d be a positive integer, U an open interval and $f : U \to \mathbb{R}$ a function with $2d+1$ derivatives. Suppose $Df(x) > 0$ for all $x \in U$. Then $S_k(f) \geq 0$ on U for all $1 \leq k \leq d$ if and only if f is matrix monotone of order $d+1$.*

Proof. Suppose first that $S_k(f) \geq 0$ for all $1 \leq k \leq d$. Take some $x \in U$ and let $\mathcal{R}$ be the dth Padé approximant to f at x. This rational map is in the Pick class because $S_k(\mathcal{R})(x) = S_k(f)(x) \geq 0$ for $1 \leq k < d$ (Lemma 8), so $M_{d+1}(x, \mathcal{R})$ is a positive matrix by [D, Theorem III.IV]. Since f and $\mathcal{R}$ coincide to order $2d$ at x and $D^{2d+1}(f)(x) = D^{2d+1}(\mathcal{R})(x) +$

$Df(x)S_d(f)(x)$ (this is the alternative definition of the Schwarzian derivative from Section 2),

(31)
$$M_{d+1}(x,f) = M_{d+1}(x,\mathcal{R}) + Df(x)S_d(f)(x)\begin{bmatrix} 0 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \geq M_{d+1}(x,\mathcal{R}).$$

Thus $M_{d+1}(x,f)$ is also a positive matrix. Since $M_{d+1}(x,f)$ is positive for every $x \in U$, Theorem VIII.V of [D] shows ([1]) that f is matrix monotone of order $d+1$.

Now suppose that f is matrix monotone of order $d+1$ on V. Choose some point $x \in U$. Applying [D, Theorem XIV.I] with $n = d+1$ and S consisting of $2d+1$ copies of x gives a Pick function ϕ on U which coincides with f to order $2d$ at x. According to Theorem III.IV of [D], either $M_d(x,\phi)$ is strictly positive, or ϕ is a rational map of degree at most $d-1$. In this last case, [D, Theorem XIV.II] states that f and the rational Pick map ϕ coincide on U; the result is then immediate from Lemma 8. So suppose $M_d(x,\phi) = M_d(x,f)$ is strictly positive. Then the principal minors of $M_d(x,f)$ are strictly positive: $\det M_j(x,f) > 0$ for $1 \leq j \leq d$. Thus f is normal of orders $1,\ldots,d$ at x and formula (8) can be freely applied. This gives $S_k(f)(x) > 0$ for $1 \leq k < d$. Furthermore, $M_{d+1}(x,f)$ is a positive matrix by [D, Theorem VII.VI], so $S_d(f)(x) \geq 0$. ∎

It is a remarkable fact (Loewner's Theorem [D]) that a function is matrix monotone of all orders if and only if it extends holomorphically to the complex upper half-plane $\mathbb{H}$ and maps $\mathbb{H}$ into itself. Thus:

Proof of Proposition 1. Combine Lemma 10 and Loewner's theorem. ∎

7. Proof of the generalized Koebe lemma. In this section we prove Theorem 2. Without loss of generality $Df > 0$ on U, so f is matrix monotone of order $d+1$ (Lemma 10). Applying [D, Theorem XIV.I] with $n = d+1$ and S consisting of $2d+1$ copies of x gives a Pick function ϕ on U which coincides with f to order $2d$ at x. It clearly suffices to prove the result for ϕ. Now ϕ, being in the Pick class, has the integral representation

$$\phi(x) = \alpha x + \beta + \int \left[\frac{1}{\xi - x} - \frac{\xi}{\xi^2+1}\right] d\mu(\xi), \tag{32}$$

([1]) The convexity hypothesis in the theorem is used to get the existence of sufficiently many derivatives almost everywhere; if the function is assumed sufficiently differentiable, as here, then this hypothesis is automatically satisfied—it follows from the positivity of the matrix.

where $\alpha \geq 0$, β is real and μ is a positive Borel measure, supported in $\mathbb{R} \setminus U$, for which $\int (\xi^2 + 1)^{-1} \, d\mu(\xi)$ is finite. See [D, Theorem II.I, Lemma II.2]. The derivatives of ϕ are given by the following formulae:

$$\begin{aligned} D\phi(x) &= \alpha + \int \frac{1}{(\xi - x)^2} \, d\mu(\xi) \\ D^m\phi(x) &= m! \int \frac{1}{(\xi - x)^{m+1}} \, d\mu(\xi) \quad \text{if } m > 1. \end{aligned} \tag{33}$$

We can now estimate $|D^m\phi(x)|$ easily:

$$\begin{aligned} |D^m\phi(x)| &= m! \left| \int \frac{1}{(\xi - x)^{m+1}} \, d\mu(\xi) \right| \\ &\leq m! \int \frac{1}{|\xi - x|^{n+1}} \frac{1}{|\xi - x|^{m-n}} \, d\mu(\xi) \\ &\leq \frac{m!}{n!} \frac{1}{\operatorname{dist}(x, \partial U)^{m-n}} |D^n\phi(x)|. \end{aligned} \tag{34}$$

The last assertion holds because n is odd, so $|\xi - x|^{n+1} = (\xi - x)^{n+1}$, and because μ puts no mass on the interval U. Also notice that in the case $n = 1$ we have used the positivity of α.

8. Proof of the main theorem. In this section we prove Theorem 1. Let us recall a few definitions. An interval is called *nice* if the iterates of its boundary points never return inside the interval. A sequence of intervals $\{W_j\}_{j=0}^{s}$ is called a *chain* if W_j is a connected component of $f^{-1}(W_{j+1})$. The *order* of a chain is the number of intervals in the chain containing a critical point. A δ-*scaled neighbourhood* of an interval J is any V containing the set $\{x : \exists y \in J, |x-y| < \delta|J|\}$. In this case we also say that J is δ-*well-inside* V. In what follows we will assume that I is the interval $[0,1]$.

We need real bounds, and will use Theorem D′ of [SV, Section 8]. The authors formulate this result slightly differently, however their proof gives precisely this:

FACT 1. *Let $f : I \to I$ be a C^3 map with non-flat critical points. Then there exists $\tau > 0$ and arbitrarily small neighbourhoods $\mathcal{W}$ and $\mathcal{V} \supset \mathcal{W}$ of the set of those critical points which are not in the basin of any periodic attractor, such that*

- *all connected components of $\mathcal{W}$ and $\mathcal{V}$ are nice intervals;*
- *if $W \subset V$ are two connected components of $\mathcal{W}$ and $\mathcal{V}$, then V is a τ-scaled neighbourhood of W;*
- *if $x \in I$ is a point and $s \geq 0$ is minimal such that $f^s(x) \in \mathcal{W}$, then the chain obtained by pulling back the connected component of $\mathcal{V}$ containing $f^s(x)$ along the orbit $x, f(x), \ldots, f^s(x)$ has order bounded by the number of critical points of f.*

We also make use of Theorem 2 from [ST], which we state as:

FACT 2. *Let $f : I \to I$ be a C^n map with non-flat critical points, $n \geq 2$. Let T be an interval such that $f^s : T \to f^s(T)$ is a diffeomorphism. For each $S, \tau, \epsilon > 0$ there exists $\delta = \delta(S, \tau, \epsilon, f) > 0$ satisfying the following. If $\sum_{j=0}^{s-1} |f^j(T)| \leq S$ and J is a subinterval of T such that*

- *$f^s(T)$ is a τ-scaled neighbourhood of $f^s(J)$,*
- *$|f^j(J)| < \delta$ for $0 \leq j < s$,*

then, letting $\phi_0 : J \to I$ and $\phi_s : f^s(J) \to I$ be affine diffeomorphisms, there exists a real-analytic diffeomorphism $G : I \to I$ such that $\|\phi_s f^s \phi_0^{-1} - G\|_{C^n} < \epsilon$, and G^{-1} belongs to the $P_\infty((-\tau/2, 1 + \tau/2))$ class.

In addition, we will use the following lemmas:

LEMMA 11. *Let $F, G : I \to I$ be two C^n diffeomorphisms, $n \geq 1$, with $\|F - G\|_{C^n} < \epsilon$. Take some $K, \epsilon > 0$ and suppose $|DG(x)| > K^{-1}$ and $|D^k G(x)| < K$ for all $x \in I$ and $k = 1, \ldots, n$. Then there exists $\delta = \delta(n, K, \epsilon)$ such that $\|F \circ G^{-1} - \mathrm{Id}\|_{C^n} < \delta$. Moreover, $\lim_{\epsilon \to 0} \delta(n, K, \epsilon) = 0$.*

Proof. Denote $H = F \circ G^{-1}$. Then $|H(x) - x| = |F - G|(G^{-1}(x)) < \epsilon$. The first derivative of H is $DH(x) = (DF/DG)(G^{-1}(x))$ so obviously $|DH(x) - 1| < \epsilon K$. This proves the lemma for $n = 1$ with $\delta(1, K, \epsilon) = \epsilon K$. For $n \geq 2$ we reason inductively. Observe that

$$(D^n H) \circ G$$
$$= \frac{D^n F - D^n G + D^n G(1 - H' \circ G) - \sum_{i=2}^{n-1} (D^i H) \circ G Q_{n,i}(DG, \ldots, D^{n+1-i} G)}{(DG)^n}$$

where the $Q_{n,i}$ are polynomials (this is easily checked by induction).

Since $\|DG\|_{C^{n-1}} < K$ by hypothesis, there is some $\Delta = \Delta(n, K)$ for which $|Q_{n,i}(DG, \ldots, D^{n+1-i} G)| \leq \Delta$ for $i = 2, \ldots, n-1$. Thus

$$\begin{aligned} |D^n H| &\leq \frac{\|F - G\|_{C^n} + \|DG\|_{C^{n-1}} \epsilon K + \sum_{i=2}^{n-1} \|H - \mathrm{Id}\|_{C^i} \Delta(n, K)}{K^{-n}} \\ &\leq K^n(\epsilon + \epsilon K^2 + n\delta(n-1, K, \epsilon)\Delta(n, K)) \equiv \delta'(n, K, \epsilon). \end{aligned} \tag{35}$$

Clearly $\delta'(n, K, \epsilon) \to 0$ as $\epsilon \to 0$. The result follows. ∎

LEMMA 12. *For every $A \geq 0$, $\alpha > 1$ and $d \geq 1$ there is some $\epsilon = \epsilon(A, \alpha, d) > 0$ with the following property. If $\phi, \psi : I \to I$ are $C^{2d+1}(I)$ diffeomorphisms, $\|\phi - \mathrm{Id}\|_{C^{2d+1}} < \epsilon$, $\|\psi - \mathrm{Id}\|_{C^{2d+1}} < \epsilon$ and $0 \leq a \leq A$, then $(\psi \circ q_{\alpha,a} \circ \phi)^{-1} \in P_d(\mathrm{int}\, I)$ where*

$$q_{\alpha,a} : x \mapsto \frac{(x + a)^\alpha - a^\alpha}{(1 + a)^\alpha - a^\alpha}. \tag{36}$$

Recall that we have assumed $I = [0, 1]$.

Proof. The only difficulty here is that $q_{\alpha,a}$ has a singularity at $-a$ which may be arbitrarily close to I. This means that derivatives of $q_{\alpha,a}^{-1}$ are not uniformly bounded as $a \to 0$, complicating continuity arguments. In what follows, we only mark dependence on a explicitly: the other parameters α, A and n should be considered as fixed, with all quantities potentially depending on them. For example, we will write q_a for $q_{\alpha,a}$.

The singularity can be side-stepped by decomposing q_a as $t \circ s_a \circ r$ where $s_a : I \to I$ is a real-analytic homeomorphism, diffeomorphic on $\operatorname{int} I$, with s_a^{-1} in the Pick class on $\operatorname{int} I$, like q_a. The maps $r, t : I \to I$ should be real-analytic diffeomorphisms, with r^{-1} and t^{-1} Pick functions on J, an interval strictly bigger than I. Finally, r^{-1} and t^{-1} should not be rational functions. Then the the matrices $M_k(y, r^{-1})$ and $M_k(y, s^{-1})$ will be uniformly positive for $y \in I$ and $k = 1, \dots, n$ thanks to [D, Theorem III.IV]. Note that r, t and J do not depend on a. Although it is not hard to give an explicit such decomposition, we will not do so here since the formulae are ugly and uninformative.

With such a decomposition in hand, [D, Theorem VII.V] and an easy continuity argument imply that $(\psi \circ t)^{-1} \in P_d(I)$ and $(r \circ \phi)^{-1} \in P_d(I)$ if $\|\phi - \mathrm{Id}\|_{C^{2d+1}}$ and $\|\psi - \mathrm{Id}\|_{C^{2d+1}}$ are sufficiently small, which is the desired result. ∎

Now we can finish the proof of Theorem. The notation $g \in P_d^{-1}(x)$ means $S_1(g^{-1})(g(x)) \geq 0, \dots, S_d(g^{-1})(g(x)) \geq 0$ (local inverse at x). Assume that a critical point c is not contained in the basin of a periodic attractor. The neighbourhood X of c will be a connected component of $\mathcal{W}$ given by Fact 1 for $\mathcal{W}$ sufficiently small.

Take $\mathcal{W}$ and $\mathcal{V}$ as in Fact 1 and let $f^s(x) \in \mathcal{W}$. We may suppose that $s \geq 0$ is minimal with $f^s(x) \in \mathcal{W}$, since the general case can be deduced from this by decomposition (recall that the class P_d is closed under composition).

Let V and W be connected components of $\mathcal{V}$ and $\mathcal{W}$ containing c. Let V' be a $\tau/3$-scaled neighbourhood of W so that $|V'| = (1 + 2/3\tau)|W|$. The interval V is also a $\tau/3$-scaled neighbourhood of V'. Let $\{V_j\}_{j=0}^s$ be the corresponding chain of pullbacks of V along the orbit of x, i.e. $f^j(x) \in V_j$ and $V_s = V$, and let $\{V'_j\}_{j=0}^s$ and $\{W_j\}_{j=0}^s$ be corresponding chains for V' and W. The map $f^s : W_0 \to W$ is a diffeomorphism because of the minimality of s. This is true even if $\mathcal{W}$ does not contain every critical point, as long as $\mathcal{W}$ is sufficiently small (the necessary smallness does not depend on x or s).

Due to [SV, Theorem C] there exist $\tau' > 0$ and $C > 0$ such that V'_j is a τ'-scaled neighbourhood of W_j, V_j is a τ'-scaled neighbourhood of V'_j and $|V'_j| < C|W_j|$ for all $j = 0, \dots, s$. Minimality of s implies that all intervals

W_j are disjoint, so

$$\sum_{j=0}^{s} |V'_j| < C$$

and we can use Fact 2 for $f^k : V'_j \to V'_{j+k}$ if this map is a diffeomorphism.

Let $0 \le s_0 < s_1 < \cdots < s_k = s$ be the moments j when V'_j contains a critical point, and put $s_{-1} = -1$ for convenience. Then $f^{s_i - s_{i-1} - 1} : V'_{s_{i-1}+1} \to V'_{s_i}$ is a diffeomorphism for $i = 0, \ldots, k$. It is enough to show that each $f^{s_i - s_{i-1}}$ belongs to $P_d^{-1}(f^{s_{i-1}+1}(x))$ for $i = 0, \ldots k$, since these compose to give f^{s+1}.

Take some $0 \le i < k$ (see below for the case $i = k$) and apply Fact 2 to $f^{s_i - s_{i-1} - 1} : V'_{s_{i-1}+1} \to V'_{s_i}$ with $T = V'_{s_{i-1}+1}$ and $J = W_{s_{i-1}+1}$. Let $F_i : I \to I$ be $f^{s_i - s_{i-1} - 1}$ pre- and post-composed with affine maps taking I to $W_{s_{i-1}+1}$ and W_{s_i} to I respectively. We obtain a diffeomorphism $G_i : I \to I$ with G_i^{-1} in the $P_\infty((-\tau/6, 1+\tau/6))$ class such that $\|F_i - G_i\|_{C^{2d+1}} < \epsilon$. The complex Koebe lemma gives the bounds on G_i needed to apply Lemma 11, yielding $\|F_i \circ {G_i}^{-1} - \mathrm{Id}\|_{C^{2d+1}} < \delta$. Note that by shrinking the neighbourhood $\mathcal{W}$ we can make ϵ and δ as small as we like.

Abusing the notation, let c be a critical point contained in V'_{s_i}. If $\mathcal{W}$ is small enough, it will be the only critical point. Note that c is not contained in W_{s_i} because we are considering the case $s_i \ne s$. Let $F : I \to I$ be $f : W_{s_i} \to f(W_{s_i})$ pre- and post-composed with affine maps taking I to W_{s_i} and $f(W_{s_i})$ to I respectively, as in the previous paragraph. Then F can be written in the form $\phi \circ q_{\alpha,a} \circ \psi$, where $q_{\alpha,a}(x) = ((x+a)^\alpha - a^\alpha)/((1+a)^\alpha - a^\alpha)$, and ϕ, ψ are diffeomorphisms of I which are close to the identity map in the C^{2d+1} topology if V'_{s_i} is small. These assertions on F follow from the definition of a critical point being non-flat.

As noted above, the intervals V'_{s_i} and W_{s_i} are comparable, so c is not far away from the interval W_{s_i} compared to its size. Expressed in terms of the rescaled map F, this means that there exists a uniform constant $A > 0$ such that the parameter a is always in $[0, A]$. Applying Lemma 12, we see that the inverse of $\phi \circ q_{\alpha,a} \circ (\psi \circ F_i \circ G_i^{-1})$ is in $P_d(I)$, at least if $\mathcal{W}$ is small enough. Composing with G_i, we see that the inverse of $\phi \circ q_{\alpha,a} \circ \psi \circ F_i$ is also in $P_d(\operatorname{int} I)$. Since this composition is precisely $f^{s_i - s_{i-1}} : W_{s_{i-1}+1} \to W_{s_i+1}$ rescaled affinely, this shows that $f^{s_i - s_{i-1}}$ belongs to $P_d^{-1}(f^{s_{i-1}+1}(x))$ as claimed.

We now consider the case $i = k$, when $s_i = s$. The difference here is that the critical point belongs to $W_s = W$, which actually makes the argument slightly simpler. The critical point cuts the interval W in half; let W' denote the half containing $f^s(x)$. Then we repeat the above argument, but instead of rescaling $f : W \to f(W)$, we rescale $f : W' \to f(W')$; the parameter a is then always zero. The argument is otherwise essentially the same.

References

[B] G. A. Baker, *Essentials of Padé Approximants*, Academic Press, New York, 1975.

[D] W. F. Donoghue, *Monotone Matrix Functions and Analytic Continuation*, Springer, Berlin, 1974.

[GS] J. Graczyk and D. Sands, *Negative Schwarzian derivative by analytic change of coordinates*, in preparation.

[K] O. S. Kozlovski, *Getting rid of the negative Schwarzian derivative condition*, Ann. of Math. 152 (2000), 743–762.

[MS] O. Martio and J. Sarvas, *Injectivity theorems in plane and space*, Ann. Acad. Sci. Fenn. Ser. A I Math. 4 (1979), 383–401.

[MvS] W. de Melo and S. van Strien, *One-Dimensional Dynamics*, Springer, Berlin, 1993.

[NS] E. M. Nikishin and V. N. Sorokin, *Rational Approximations and Orthogonality*, Transl. Math. Monogr. 92, Amer. Math. Soc., Providence, RI, 1991; Russian original: Nauka, Moscow, 1988.

[P] H. Padé, *Sur la représentation approchée d'une fonction par des fractions rationnelles*, Ann. Sci. École Norm. Sup. (3) 9 (1892), Suppl., 3–93.

[ST] W. X. Shen and M. Todd, *Real C^k Koebe principle*, Fund. Math. 185 (2005), 61–69.

[SV] S. van Strien and E. Vargas, *Real bounds, ergodicity and negative Schwarzian for multimodal maps*, J. Amer. Math. Soc. 17 (2004), 749–782; erratum, ibid. 20 (2007), 267–268.

[T] W. P. Thurston, *Zippers and univalent functions*, in: The Bieberbach Conjecture (West Lafayette, IN, 1985), Math. Surveys Monogr. 21, Amer. Math. Soc., Providence, RI, 1986, 185–197.

O. Kozlovski
Mathematics Institute
University of Warwick
Coventry CV4 7AL, United Kingdom
E-mail: oleg@maths.warwick.ac.uk

D. Sands
CNRS, Département de Mathématiques
Université Paris-Sud
91405 Orsay Cedex, France
E-mail: duncan.sands@math.u-psud.fr

Received 2 January 2009;
in revised form 13 May 2009